

Fri Jun 4 10:08:27 1999

US08945574.pep

Sequence 1, Application US/08945574
GENERAL INFORMATION:
APPLICANT: Lenting, Hermanus Bernardus Maria
APPLICANT: Van Beckhoven, Rudolf Franciscus Wilhelmus Cornelis
APPLICANT: Maurer, Karl-Heinz
APPLICANT: Kottwitz, Beatrix
APPLICANT: Weiss, Albrecht
APPLICANT: Van Solingen, Pieter
TITLE OF INVENTION: Detergents Comprising Cellulases
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Henkel Corporation
STREET: 140 Germantown Pike, Suite 150
CITY: Plymouth Meeting
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19462
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: MS Word 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,574
FILING DATE: unavailable
CLASSIFICATION: C11D 3/386
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/01755
FILING DATE: 26 Apr. 1996
APPLICATION NUMBER: EP 95201115.3
FILING DATE: 28 Apr. 1995
APPLICATION NUMBER: U.S. 614,115
FILING DATE: 12 Mar. 1996
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Glenn E. J.
REGISTRATION NUMBER: 33,539
REFERENCE/DOCKET NUMBER: H 1920 PCT/US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 832-2228
TELEFAX: (610) 941-6067
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 457 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-945-574-1
MKRTTIFAVLLMTLALFSIGNTTAADDYSVVEHGLSISNGELVNERGEYVQLKGMSSHGLQWYGFV
NYESMKLWDDMGITVPEAAYTSSTGGYIDDPVSKEVKEVTEAIDLGIVLIDHLLSDNDPHIYKEE
AKDFDEMSLGYDPVNYVILANEPNGSYTDNCKIPYAEVPIVIRNDNNVIVGTGWSQDVHH
AANDQLADPNVMAHFTAGTGNLQADQVDYALOGGAIFVSENGTSAAATGGGVFLDEAQWIDFMDH
RNLNANWLSLTHKDESSAALMPGANPTGGWTEAELSPSGTFVREKIREASIPSPDTPPSGPDPE
POPTPPSPGEPYPAWDSNQIYTNIEVYHNGOLWQAKWWTQNOEPDYGWPEPLNI
PGEYPAWDSNQIYTNIEVYHNGOLWQAKWWTQNOEPDYGWPEPLNI

Sequence 2, Application US/08945574
GENERAL INFORMATION:
APPLICANT: Lenting, Hermanus Bernardus Maria
APPLICANT: Van Beckhoven, Rudolf Franciscus Wilhelmus Cornelis
APPLICANT: Maurer, Karl-Heinz
APPLICANT: Kottwitz, Beatrix
APPLICANT: Weiss, Albrecht
APPLICANT: Van Solingen, Pieter
TITLE OF INVENTION: Detergents Comprising Cellulases
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Henkel Corporation
STREET: 140 Germantown Pike, Suite 150
CITY: Plymouth Meeting
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19462
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: MS Word 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,574
FILING DATE: unavailable
CLASSIFICATION: C11D 3/386
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/01755
FILING DATE: 26 Apr. 1996
APPLICATION NUMBER: EP 95201115.3
FILING DATE: 28 Apr. 1995
APPLICATION NUMBER: U.S. 614,115
FILING DATE: 12 Mar. 1996
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Glenn E. J.
REGISTRATION NUMBER: 33,539
REFERENCE/DOCKET NUMBER: H 1920 PCT/US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 832-2228
TELEFAX: (610) 941-6067
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 574 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-945-574-2
MKHKSVMWLVVYVSVFAPVSSANEDVKTLIDIOSYVRDMQPCWNLGNTFDVAGQDETAMGNPRVTRH
LIERADGTSKIRIPVTEWRIGGADYIDPQFLNRVDEYVOWALEEDLYVNLHDSWLIYEMEH
NYNGVAKTSVNEQSLSHFXYDTKLFDSVNEPKFSQNGEIRHNHALLDLDLNTVFVIVROSGGON
DIRPLVPTMETATSOPLNNLYOTIDKDDPNLIATVHYGFWPFSVNIAGYTRFEEDSKREIIFEDR
VHHTFVARGIPVYVGLGFLGDKHTGVIQOGKFAFFELIHLINERDITHMLWNGOHFNHTYKVD
BELFDMLRASWGRSSVAESNFIYLKQGDRIADATVTLQHGNETJGLQANGRLTFQGDYELNGRLIV
KARVLSAAGSGTLATNGWTAENRGAHFRVNTYRTPVLSQTOGHVSNFSPISAFNGNSLATNEAVY
VDGNGRPGDWTSEFEGYAFSPSIDTHEIRLTEAFFREVROGEVRLTFHFWSGEIVNYTIKNGNQVVG
IAQTINSNNKNNKI

MPREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

mparch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 4 09:55:50 1999; MasPar time 14.37 Seconds
Tabular output not generated. 918.961 Million cell updates/sec

Title: >US-08-945-574-1
Description: (1-467) from US08945574.pep
Perfect Score: 3419
Sequence: 1 MKKITTIFAVLLMTALFSI.....KWTQNBFGDPYGPWEPLN 467

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 50.573; Variance 100.476; scale 0.503

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2828	82.7	409	1	GUN2_BACS4 ENDOGLUCANASE B (EC 3.1.1.1)	0.00e+00
2	2803	82.0	488	1	GUN1_BACS4 ENDOGLUCANASE A (EC 3.1.1.1)	0.00e+00
3	1664	48.7	505	1	GUNV_ERWCA ENDOGLUCANASE V PRECURSOR	0.00e+00
4	1660	48.6	499	1	GUN1_BACSU ENDOGLUCANASE PRECURSOR	0.00e+00
5	1659	48.5	444	1	GUNN_ERWCA ENDOGLUCANASE N PRECURSOR	0.00e+00
6	1641	48.0	499	1	GUN2_BACSU ENDOGLUCANASE PRECURSOR	0.00e+00
7	1616	47.3	504	1	GUNW_ERWCA ENDOGLUCANASE V1 PRECURSOR	2.51e-302
8	1598	46.7	499	1	GUN3_BACSU ENDOGLUCANASE PRECURSOR	1.64e-298
9	1491	43.6	448	1	GUN3_CLOAB ENDOGLUCANASE PRECURSOR	7.41e-276
10	884	25.9	429	1	GUNA_BTUFI ENDOGLUCANASE A PRECURSOR	1.04e-148
11	816	23.9	426	1	GUN2_ERWCH ENDOGLUCANASE Z PRECURSOR	1.08e-134
12	810	23.7	312	1	GUN4_RUMAL ENDOGLUCANASE IV (EC 3.1.1.1)	1.84e-133
13	805	23.5	466	1	GUN5_THERU ENDOGLUCANASE E-5 PRECURSOR	1.96e-132
14	739	21.6	800	1	GUN_BACSI ENDOGLUCANASE PRECURSOR	6.52e-119
15	727	21.3	459	1	GUNA_STRLI ENDOGLUCANASE CELA PRECURSOR	1.84e-116
16	690	20.2	941	1	GUN_BACS6 ENDOGLUCANASE PRECURSOR	6.29e-109
17	613	17.9	825	1	GUN3_BACS4 ENDOGLUCANASE C PRECURSOR	2.39e-93
18	186	5.4	848	1	GUND_CLOCE ENDOGLUCANASE D PRECURSOR	1.40e-12
19	179	5.2	900	1	GUNH_CLOTHM ENDOGLUCANASE H PRECURSOR	1.88e-11
20	175	5.1	820	1	CHIA_ALRSO CHITININASE A PRECURSOR	8.17e-11
21	139	4.1	826	1	SPF2_PLAYO SPOROZOITE SURFACE PRO	2.39e-05
22	128	3.7	897	1	YHEB_ECOLI HYPOTHETICAL 97.1 KD P	8.45e-04
23	122	3.6	699	1	VGLG_HSV2H GLYCOPROTEIN G.	5.51e-03

24	124	3.6	1331	1	MANB_CALSA	BETA-MANNANASE / ENDOS	2.97e-03
25	120	3.5	316	1	UNC_PVKA	URACIL-DNA GLYCOSYLASE	1.02e-02
26	116	3.4	339	1	CSP_PLABE	CIRCUMSPOROZOITE PROTE	3.40e-02
27	115	3.4	347	1	CSP_PLABA	CIRCUMSPOROZOITE PROTE	4.57e-02
28	116	3.4	353	1	CA26_HUMAN	COLLAGEN ALPHA 2(VI) C	3.40e-02
29	117	3.4	449	1	GUNA_CELFI	ENDOGLUCANASE A PRECUR	3.40e-02
30	116	3.4	825	1	5E5_RAT	5E5 ANTIGEN	2.52e-02
31	112	3.3	456	1	GUNA_MICBI	ENDOGLUCANASE A PRECUR	1.10e-01
32	114	3.3	487	1	EBN2_EBV	EBNA-2 NUCLEAR PROTEIN	6.14e-02
33	112	3.3	658	1	GUN3_FIBSU	ENDOGLUCANASE 3 PRECUR	1.10e-01
34	108	3.2	143	1	PARI_TRYBB	PROCYCLIC FORM SPECIFI	3.49e-01
35	108	3.2	145	1	PARC_TRYBB	PROCYCLIC FORM SPECIFI	3.49e-01
36	108	3.2	185	1	ADML_HUMAN	ADRENOMEDULLIN PRECURS	3.49e-01
37	108	3.2	295	1	LEG3_CANFA	GALECTIN-3 (GALACTOSE-	3.49e-01
38	108	3.2	315	1	UNG_PRVIF	URACIL-DNA GLYCOSYLASE	3.49e-01
39	108	3.2	349	1	FLAV_SOLTU	FLAVONOL SYNTHASE (EC	3.49e-01
40	109	3.2	484	1	GUX_CELFI	EXOGLUCANASE PRECURSOR	2.62e-01
41	111	3.2	917	1	SVI_STAAN	ISOLEUCYL-TRNA SYNTHET	1.48e-01
42	111	3.2	1029	1	CA26_MOUSE	COLLAGEN ALPHA 2(VI) C	1.48e-01
43	110	3.2	1664	1	SLPI_CLOTM	CELL SURFACE GLYCOPROT	1.97e-01
44	106	3.1	459	1	G33_RAT	GENE 33 POLYPEPTIDE.	6.12e-01
45	107	3.1	1018	1	FNBA_STAAN	FIBRONECTIN-BINDING PR	4.63e-01

ALIGNMENTS

RESULT 1

ID GUN2_BACS4 STANDARD; PRT; 409 AA.

AC P06565;

DT 01-JAN-1988 (REL. 06, CREATED)

DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)

DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)

DE ENDOGLUCANASE B (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE)

DE (CLONE PNK2).

GN CELB.

OS BACILLUS SP. (STRAIN N-4).

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;

OC BACILLUS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 8705694.

RA FUKUMORI F., SASHIHARA N., KUDO T., HORKOSHI K.;

RT "Nucleotide sequences of two cellulase genes from alkalophilic

RT Bacillus sp. strain N-4 and their strong homology.";

RL J. BACTERIOL. 168:479-485(1986).

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC

CC LINKAGES IN CELLULOSE.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL

CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; M14729; G142656; -

DR PIR; B25156; B25156.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

DR PFAM; PF00150; cellulase; 1

CC CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE.

FT ACT_SITE 165 165 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 254 254 NUCLEOPHILE (BY SIMILARITY).

SQ SEQUENCE 409 AA; 45690 MW; C77CF2F8 CRC32;

Query Match 82.7%; Score 2828; DB 1; Length 409;

Best Local Similarity 94.6%; Pred. No. 0.00e+00;

Matches 387; Conservative 12; Mismatches 6; Indels 4; Gaps 2;

Db 1 MKKITTIFAVLLMTALFIIGNTTAADDYVVEEHGQLSISNGELVNDRGEPVQLKGMS 60

||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1 MKKLTTFIVFTLAL-LF-VGNSTSANNGSVVEQNGOLSIQNGOLVNEHGDVPOLKGMSS 58

```

RESULT 3
ID  GENV_ERWCA  STANDARD;  PRT;  505  AA.
AC  Q47096;
DT  01-NOV-1997 (REL. 35, CREATED)
DT  01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT  01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE  ENDOGLUCANASE V PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V)
DE  (CELLULOSE V).
GN  CELV.
OS  ERWINIA CAROTOVORA.
OC  BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC  ERWINIA.
CC  [1]
CC  SEQUENCE FROM N.A.
CC  STRAIN=SCRI193;
CC  MEDLINE: 94067016.
CC  COOPER V.J.C.; SALMOND G.P.C.;
RA  "Molecular analysis of the major cellulase (Celv) of Erwinia
RT  carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT  domains.";
RL  MOL. GEN. GENET. 241:341-350(1993).
CC  -1- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH
CC  OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES
CC  CELSIUS.
CC  -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC  LINKAGES IN CELLULOSE.
CC  -1- SUBCELLULAR LOCATION: SECRETED.
CC  -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC  HYDROLASES).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collabora-
CC  tion between the Swiss Institute of Bioinformatics and the EMBL outstat-
CC  the European Bioinformatics Institute. There are no restrictions on
CC  use by non-profit institutions as long as its content is in no
CC  modified and this statement is not removed. Usage by and for comm-
CC  enties requires a license agreement (See http://www.isb-sib.ch/announc
CC  or send an email to license@sib-sib.ch).

```

[illegible]

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L39788; G662361; -
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00942; CBD 3; 1.
KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 444 ENDOGLUCANASE N.
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 444 AA; 48300 MW; 6CA14F71 CRC32;

Query Match 48.5%; Score 1659; DB 1; Length 444;
Best Local Similarity 67.6%; Pred. No. 0.00e+00;
Matches 217; Conservative 51; Mismatches 50; Indels 3; Gaps 3;

Db 35 VETHQGLSTENGRLVDEQCKRVQLRGVSHGLQWFGDYVYKDSMKWLRRDWDGIVNFRVAM 94
Y ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
32 VEEHQGLSISNGELVNERGEQVQLGMSHGLQYGFVYVESMKWLRRDWDGIVFRAM 91
D ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 YTAADGYISNPSLANKVKAVAAGSLGVIYIIDHILSDNDPNYKAAKTFEAMAGL 154
D ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
92 YTSGGYIDDPVKEVKETVZAAADLGIYIIDHILSDNDPNYKEAKOFFDENSEL 151
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 YGSSPNVIYEIANEPNGG-VTWNGQIRYALEYDTDIRSKDPNLIIIVGTWSDIHDA 213
D ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 YGDYENVIYEIANEPNGSDVTWQNDQIKPYAEVPIVRDNDPNIIIVGTWSDVHHA 211
D 214 ADNQLPDNTLYALHFYAGTHQGLRDRIDYAGSRGAAIFVSEWGTSDASNGGPELPS 273
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
212 ADNLDADPNVYAFYFAGTHQGLRDRIDYAGSRGAAIFVSEWGTSAATGDGVFLDEA 271
D 274 QTWIDFLNRRGVSNWVSTDKSEASALPGAKSGGWTQNLSTSGRFVREIRAGAN 333
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 QWIDFMEERNLSWANSVTHKDESSAALPGANPTGGWTEAELSPGTFVREKIRASAS 331
D 334 LGGSDTPTPTPTPTPTPTNPN 354
D ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
332 IPPSD-PTPPSDGPB-DPGE 350

RESULT 6
ID GUN2_BACSU STANDARD; PRT; 499 AA.
AC P10475;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMCASE) (CELLULOSE).
GN BGL OR GLD OR EGLS.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PAP115;
RX MEDLINE; 87066783.
RA MACKAY R.M., LO A., WILLOCK G., ZUKER M., BAIRD S., DOVE M.,
RA MORANELLI F., SELIGY V.;
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL NUCLEIC ACIDS RES. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CK-2;
RX MEDLINE; 95225656.

```

```

RA LINDAHL V., AA K., TRONSMO A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RT subtilis CK-2.";
RL ANTONIE VAN LEEUWENHOEK 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA ROSE M., ENTIAN K.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE OF 30-45.
RC STRAIN=CK-2;
RX MEDLINE; 95225655.
RA AA K., FLENGSROD R., LINDAHL V., TRONSMO A.;
RT "Characterization of production and enzyme properties of an
RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RT compost soil.";
RL ANTONIE VAN LEEUWENHOEK 66:319-326(1994).
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29076; G509267; -
DR EMBL; X04689; G39824; -
DR EMBL; X67044; G39777; -
DR EMBL; Z73234; E249661; ALT_INIT.
DR PIR; A26114; A26114.
DR SUBTILIST; BG10437; BGLC.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00942; CBD 3; 1.
DR HSP; 006851; NBC.
KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
FT SIGNAL 1 29
FT CHAIN 30 499 ENDOGLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT CONFLICT 283 283 S -> N (IN REF. 2).
SQ SEQUENCE 499 AA; 55287 MW; 1DE2AA90 CRC32;

Query Match 48.0%; Score 1641; DB 1; Length 499;
Best Local Similarity 62.7%; Pred. No. 0.00e+00;
Matches 222; Conservative 60; Mismatches 66; Indels 6; Gaps 6;

Db 1 MKRSISIFITCLITLLTMGMIA SPASAGTKTPVAKNGQLSIKGTQLVNRDKAVQLK 60
Y ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1 MKKITTIFAV-LLMTL-ALFS-IGNTTAADYSV-VEEHQGLSISNGELVNERGEQVQLK 56
D 61 GTSSEGLQWGEYVKNKSLKWLRRDWDGIVFRAMTADGGYIDNFSVKNKVEAEAAK 120
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 GMSSEGLQWGEYVKNKSLKWLRRDWDGIVFRAMTADGGYIDNFSVKNKVEAEAAI 116
D 121 ELGIYVIDWHIINDGNPNONKEKKEFEKSSLSGNTPNVIYEIANEPNG-DVNWKR 179
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 DUGIYVIDWHIINDGNPNIIYKEEAKDFESEMSELGDPNVIYEIANEPNGSDVTWQ 176
D 180 IKPYAEVIVIRKNDPNIIIVGTWSDVNDADQDKANVMYALHFYAGTHQGL 239
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
177 IKPYAEVIVIRKNDPNIIIVGTWSDVNDADQDKANVMYALHFYAGTHQGL 236
D 240 RDKANYALSKGAPFVTEWGTSDASNGGVFLDQSEWLYLDSKTSIWNWNLSDQES 299
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 RDQVDYALDQGAIFVSEWGTSAATGDGVFLDEA QWIDFMEERNLSWANSVTHKDES 296

```

Db 300 SSALPKGASKTGGNRLSDLSASGTFVRENILGTDKDTHDIPETSKDKPTQENG 353
 QY 297 SAALMPGANPTGGTAEALSPSGTFVREKIREASIPSPDP-TPPSDGPDPG 349

RESULT 7
 ID GUNW ERWCA STANDARD; PRT; 504 AA.
 AC Q93395;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE ENDOGLUCANASE VI PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE VI)
 DE (CELLULOSE VI).

OC ERWINIA CAROTOVORA.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC ERWINIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SCC3193;
 RX MEDLINE; 95231512.
 RA MAE A., HEIKINHEIMO R., PALVA E.T.;
 RT "Structure and regulation of the Erwinia carotovora subspecies
 RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
 RT phytopathogenicity";
 RT MOL. GEN. GENET. 247:17-26(1995).
 RL CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X79241; G493493; -
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PFAM; PF00942; CBD_3; 1.
 DR HSP; Q06851; INEC
 KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
 FT SIGNAL 1 31
 FT CHAIN 32 504
 FT DOMAIN 32 334
 FT CATALYTIC.
 FT LINKER.
 FT DOMAIN 335 352
 FT DOMAIN 353 504
 FT CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 168 168
 FT ACT_SITE 256 256
 FT NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 504 AA; 54963 MW; 60AFD878 CRC32;

Query Match 47.3%; Score 1616; DB 1; Length 504;
 Best Local Similarity 65.8%; Pred. No. 2.51e-302;
 Matches 210; Conservative 55; Mismatches 52; Indels 2; Gaps 2;

Db 35 VETHGQLSIENGLRLVDEQKRVQLRGISNGLQWGYDVYVNRKMSKWLRRDGGINFRVAM 94
 QY 32 VEERHQLSINGELVNERGEQVQLKGMSSHGLQWGYQFVNVESKWLRRDGGITVFRAM 91
 Db 95 YTAENGIANPLANKVEAAAGLGVYIIDWHITLSDNDPNYKQAQKIFFAEMAGL 154
 QY 92 YTSGGYIDDPKSVKEKVKETVEAIDLGIYVIDWHIISDNDPNYKEEAKDFDEMSEL 151
 Db 155 YGNSPNVIYEIANPNSG-VTWNGOIRPYALEVDTTIRSKPDNLIIYVSGTWSODIHA 213
 QY 152 YGDPNVIYEIANPNSGSDVTWQKQIAEVIPTVNDPNNIVIVGTGWSODVHA 211
 Db 214 ADNOLPDNTLYALHFYAGTHGQFLRDRIDYAQSGAAIFVSEWGTSDASNGSGPFLPES 273

QY 212 ADNQADFNVMYAFHYAGTHGQNLRDQVDYALDQGAALFVSEWGTSAATGGVFLDEA 271
 Db 274 QTWIDFLNRRGISWYNWLSRSETSAALVAGASGGWTEQNLTSGKFVREIQRAGAG 333
 QY 272 QVWIDFMDERNLSWANWSLTHKDESSAALMPCANPTGGTAEALSPSGTFVREKIREAS 331

Db 334 LSGGDTPTMPTPTNGNG 352
 QY 332 IPPSD-PTPPSDPGEPPG 349

RESULT 8
 ID GUN3_BACSU STANDARD; PRT; 499 AA.
 AC P23549;
 DT 01-NOV-1991 (REL. 20, CREATED)
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CARBOXYMETHYL-CELLULOSE) (CMCASE) (CELLULOSE).
 GN BGIC.

OS BACILLUS SUBTILIS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 OC BACILLUS.

RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-BSE616;

RX MEDLINE; 91299280.

RA PARK S.H., KIM H.K., PACK M.Y.;

RT "Characterization and structure of the cellulase gene of Bacillus

RT subtilis BSE616";

RL AGRIC. BIOL. CHEM. 55:441-448(1991).

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC

CC LINKAGES IN CELLULOSE.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; D01057; G216388; -

DR FIR; JN0111; JN0111.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

DR PFAM; PF00150; cellulase; 1.

DR PFAM; PF00942; CBD_3; 1.

DR HSP; Q06851; INEC

KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.

FT SIGNAL 1 29

FT CHAIN 30 499

FT ACT_SITE 169 169

FT ACT_SITE 257 257

FT DOMAIN 350 499

FT SEQUENCE 499 AA; 55169 MW; 4391DCA5 CRC32;

Query Match 46.7%; Score 1598; DB 1; Length 499;
 Best Local Similarity 61.6%; Pred. No. 1.64e-298;
 Matches 218; Conservative 62; Mismatches 68; Indels 6; Gaps 6;

Db 1 MKRSISIFTCLLITLLTMGMGLASPASAGTKTPVAKNGQLSIGKTQLVNRDGRVQLK 60

QY 1 MKKTIIFAV-LLMTL-ALFSTI-GNTTADDYSV-VEEHGQLSISNGLVNERGEQVQLK 56

Db 61 GTSSGLQWGYGVNKSILKWLRRDGGITVFRAMYTADGGIIDNPVSKNKKAEVAAK 120

QY 57 GMSHGLQWGYGVNKSILKWLRRDGGITVFRAMYTADGGIIDNPVSKNKKAEVAAK 116

Db 121 ELGIYVIDWHIILNDGNPNQKREKAEFFKEMSSLYGNTPNVIYEIANPNSG-DVNWKR 179

```

QY 117 DLGIYVIDHILSDNDPNYKEAKDFDEMSELYGDPNVIYEIANEPNGSDVTWQ 176
Db 180 IKPAEAEVIRKNDPNIIIVGTQSWQDVNDRAADQOLKANDYDALHFFYAGTHGQFL 239
QY 177 IKPAEAEVIRKNDPNIIIVGTQSWQDVNDRAADQOLKANDYDALHFFYAGTHGQFL 236
Db 240 RDKANYALSAGPFIIVTGWGTSASNGGVFLDQSRWLYKLDSTKISWNNLSDKQES 299
QY 237 RDQVDYALDQGAALFVSEWGTSAATGQGVFLDEAQWIDFMDERNLSWANWSLTHKDES 296
Db 300 SSALKPGASKTGGWRLSDLSASGTFRVRENILGTDKSTKIDIPAKDKPTQENG 353
QY 297 SAALMPGANPTGGTAEALSFGTFVREKIREASIPSDP-TTPSDGPEPDPG 349

RESULT 9
ID GUN_CLOAB STANDARD; PRT; 448 AA.
AC P15704;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).
GN EGA.
OS CLOSTRIDIUM ACETOBUTYLICUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-P262;
RX MEDLINE; 88268074.
RA ZAPPE H., JONES W.A., JONES D.T., WOODS D.R.;
RT "Structure of an endo-beta-1,4-glucanase gene from Clostridium
RT acetobutylicum P262 showing homology with endoglucanase genes from
RT Bacillus spp.";
RL APPL. ENVIRON. MICROBIOL. 54:1289-1292(1988).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- THE C-TERMINAL REGION OF C.ACETOBUTYLICUM IS NOT REQUIRED FOR
CC ACTIVITY.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31311; G144790; -.
CC PIR; A27631; A27631.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC PFAM; PF00150; cellulase; 1.
CC CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
CC SIGNAL 1 34
CC CHAIN 35 448 ENDOGLUCANASE.
CC ACT_SITE 175 175 PROTON DONOR (BY SIMILARITY).
CC ACT_SITE 263 263 NUCLEOPHILE (BY SIMILARITY).
CC SEQUENCE 448 AA; 49366 MW; 5EE996E6 CRC32;

Query Match 43.6%; Score 1491; DB 1; Length 448;
Best Local Similarity 54.9%; Pred. No. 7.41e-276;
Matches 197; Conservative 69; Mismatches 88; Indels 5; Gaps 5;

Db 11 FKKTFSFLIAYVMMFTVLGNTYKAEATTSFGQGLKVQGLSDNSGKPIQLKGMSSH 70
QY 4 ITTIFAVLL-MTALFSI-GNTTAADDYVVEEHGQLSNGELVNERGEQVQLKGMSSH 61
Db 71 GLOWYVFNVDYDKFLDKGKGVNIRAMYNTEGGYISNPSQKEKIKKIYQDAIDLNM 130
QY 62 GLOWYGFVNVYESKWLKDDWGITVFRAMYTSSGGYIDDPVSV-REKVKETVEAIDLGI 120

```

```

Db 131 VYIIDHILSDNDPNYKEAKSFFQEMAEYKYSNVIYECNPNNGG-TNWANDIKPY 189
QY 121 VYIIDHILSDNDPNYKEAKSFFQEMAEYKYSNVIYECNPNNGGSDVTWQNIKPY 180
Db 190 ANYIIPAIRADPNIIIVGTQSWQDVNDRAADQOLKANDYDALHFFYAGTHGQFL 249
QY 181 AEEVPIVRNDPNIIIVGTQSWQDVNDRAADQOLKANDYDALHFFYAGTHGQFL 240
Db 250 NYAMSKGIAFVTEWGTSDASNGGVFLDESQKWDVFMASKNISMWNLALCKSKSEAAAL 309
QY 241 DYALDQGAALFVSEWGTSAATGQGVFLDEAQWIDFMDERNLSWANWSLTHKDESSAAL 300
Db 310 KSGSTTGGWTDLSLTSLFVKKSGSNTTSQTSAPTFSLQSGTYSDAQVITLTSSD 368
QY 301 MPGANPTGGTAEALSFGTFVREKIREASIPSDP-TTPSDGPEPDPGPTTPSD 358

RESULT 10
ID GUN_BUTFI STANDARD; PRT; 429 AA.
AC P22541;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE A PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE) (EGA).
GN CELA.
OS BUTYRIVIBRIO FIBRISOLVENS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC BUTYRIVIBRIO.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-50.
RC STRAIN-A46;
RX MEDLINE; 91100957.
RA HAZLEWOOD G.P., DAVIDSON K., LAURIE J.I., ROMANIEC M.P.M.,
RA GILBERT H.J.;
RT "Cloning and sequencing of the cels gene encoding endoglucanase A of
RT Butyrivibrio fibrisolvens strain A46.";
RL J. GEN. MICROBIOL. 136:2089-2097(1990).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M37031; G144155; -.
CC PIR; S29044; S29044.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC PFAM; PF00150; cellulase; 1.
CC CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
CC SIGNAL 1 34
CC CHAIN 35 429 ENDOGLUCANASE A.
CC ACT_SITE 249 249 PROTON DONOR (BY SIMILARITY).
CC ACT_SITE 334 334 NUCLEOPHILE (BY SIMILARITY).
CC SEQUENCE 429 AA; 48858 MW; 0FAFC840 CRC32;

Query Match 25.9%; Score 884; DB 1; Length 429;
Best Local Similarity 43.7%; Pred. No. 1.04e-148;
Matches 131; Conservative 65; Mismatches 93; Indels 11; Gaps 8;

Db 114 GKLAEGSHLVADGDHEVLLMGVSTHGINVPEYASAEIKSLRTGTWGNVIRLAMYTS 173
QY 36 GQLSISNGELVNERGEQVQLKGMSSHGLOWYGFVNVYESKWLKDDWGITVFRAMYTSS 95
Db 174 YNGYCVACKENQEKLDIDDAVEAATDNDYVIIDWHTLNDADPNYKADAIQFGEVY 233

```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; Y00540; G41092; -.
DR PIR; S03767; S03767.
DR PDB; IAIW; 06-MAY-98.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PFAM; PF00150; cellulase; 1.
KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL; 3D-STRUCTURE.
FT SIGNAL 1 43
FT CHAIN 44 426 ENDOGLUCANASE Z.
FT DOMAIN 44 332 CATALYTIC.
FT DOMAIN 333 366 LINKER.
FT DOMAIN 367 426 CELLULOSE-BINDING.
FT ACT_SITE 176 176 PROTON DONOR.
FT ACT_SITE 263 263 NUCLEOPHILE (BY SIMILARITY).
FT MUTAGEN 141 141 H->A: LOSS OF ACTIVITY.
FT MUTAGEN 176 176 E->A: LOSS OF ACTIVITY.
FT DISULFID 368 425
FT CONFLICT 293 295
FT CONFLICT 350 364
FT CONFLICT 388 426
FT FT
FT FT
FT FT
SQ SEQUENCE 426 AA; 46418 MW; FD9E8988 CRC32;
Query Match 23.9%; Score 816; DB 1; Length 426;
Best Local Similarity 40.8%; Pred. No. 1,08e-134;
Matches 127; Conservative 67; Mismatches 106; Indels 11; Gaps 9;
Db 76 WGGKFKFYTDATVASLKDKWKSSIVRAAGVQESGGYLDQPAGNKAKEVERVVDAIAANDMY 135
Qy | | : : : : : | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : : : : : | | | | | | | | | | | | | | | | | | | | | | | | | | |
65 WYGQ-FVNYESMKWLDDWGITVFRAAM-YTSSGGYIDDPVS-REKVETVEAAIDLGIY 121
Db 136 AIIQWHSHSAEN-N-RSEAIRFQEWMARKGNPNVIYEIYNPL--QVSWSTIKPYA 190
Qy | | : : : : : | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : : : : : | | | | | | | | | | | | | | | | | | | | | | | | | | |
122 VIIDWHILSDNDPNRYKEEKDFDEMSELGYDPNVYIEIANEPNGSDVTWDNQIKPYA 181
Db 191 EAVISAIRAPDNLIIIVGTFSWSONDEASRDIPINAKNIATYTLHFYAGTHGESLRNKAR 250
Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
182 EEVIPVINDPNPNIIVIGTGTSQDVHHADNQLADPNVMYAFHFYAGTHGQNLRQVD 241
Db 251 QALANGIALFVTEGTVNADNGGVNGVNETDAWYTFMRDNNSISNAWLNKDNEGASTYY 310
Qy | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
242 YALDQGAIFVSEWGTSAATGDGGVFLDEAQVWIDEWDELNLSWANWSLTHKDESSAALM 301
Db 311 PSDKNLTESKKVKSIQSWPYKAGSAASTDSPTDTTDTTVDPEPTTTTPATAACAN 370
Qy | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
302 PGA-NPT-GGWTEAE LSPSGTFVREKIRESAISIPPSPDPPGPGE-PDPTPPSD 358
Db 371 ANVPYPNVSKD 381
Qy : | | | | : :
359 PGEPAWDSNQ 369
RESULT 12
ID GUN4_RUMAL STANDARD; PRY; 312 AA.
AC Q07940;
DT 01-FEB-1995 (REL. 31, CREATED)
DI 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE IV (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE)
DE (EG-IV).
GN EG IV.
OS RUMINOCOCCUS ALBUS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC RUMINOCOCCUS.
RN [1]
```


SEQUENCE FROM N.A., AND SEQUENCE OF 1-19.

RP STRAIN-F-40;

RC KARIITA S., MORIOKA K., KAJINO T., SAKKA K., SHIMADA K., OHMIYA K.;

RA "Cloning and sequencing of a novel endo-1,4-beta-glucanase gene from

RT *Ruminococcus albus*.";

RL J. FERMENT. BIOENG. 76:439-444(1993).

CC -1- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF 1,4-BETA-D-GLUCOSIDIC

CC LINKAGES IN CELLULOSE.

CC -1- THIS ENZYME IS MOST ACTIVE AT PH 7 AND 40 DEGREES CELSIUS.

CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL

CC HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed, usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: D16315; G303887; --

CC DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

CC DR PFAM: PF00150; cellulase; 1

CC DR CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE.

CC KW ACT_SITE 135 135 PROTON DONOR (BY SIMILARITY).

CC FT ACT_SITE 222 222 NUCLEOPHILE (BY SIMILARITY).

CC BQ SEQUENCE 312 AA; 35766 MW; C27BDC68 CRC32;

Query Match 23.7%; Score 810; DB 1; Length 312;

Best Local Similarity 38.8%; Pred. No. 1.84e-133;

Matches 113; Conservative 69; Mismatches 100; Indels 9; Gaps 9;

Db 5 LKVNKGKLTAG-EKRPVFLGLSTHGIAWYPIECESFNALKDKWRNCIRIAMYTDEFR 63

Qy 38 LSIANGELVNERGEOVQLKGNSSHGLQWYGQFVYVESKWLRRDDWIGTVFRAAMYTSS-G 96

Db 64 GYCKDGN-KOHLKELIEKGVVIAEKLDWYIVDWHVLCDDQPMKYIDAEAEFFSDMSKRF 122

Qy 97 GYIDPSPVKVKETVE-AA-I-D-LGIYVIDWHILSDNDPNIVYKEAKDFDEMSELY 152

Db 123 ANKTNVYIEICNEPCSG-TWDK-ITEYADRIIIRSNPDALIVGTGTSTWSQDIHCA 180

Qy 153 GDYPNVYIETANEPGSDVTWQNDQIKPYAEVIVIRDNDPNIVGTGTWSQDVHAA 212

Db 181 EKPLKWDNVNLSHFYAATHGRKLSRLERICIEAGLPVFINFENLCAAGSGGDIDIDEAN 240

Qy 213 DNQLADPNVYAFHYAGTHGQNLRDQVDYALDQGAALFVSEWGTSAATGGGVFLDEAQ 272

Db 241 AWYEVIRDLGUSICSWCLNSGSDTCGVFTQNTKLSGWTDEDIKTSKIKK 291

Qy 273 VVIDFMDERNLSWANWSLTHKDESSAALMPCANPTGGWTEAELSPSGTFFVR 323

RESULT 13

GUNS_3THEFU STANDARD; PRT; 466 AA.

Q01786;

DT 01-JUL-1993 (REL. 26, CREATED)

DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE ENDOGLUCANASE E-5 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE E-4)

DE (CELLULOSE E-5) (CELLULOSE E5).

GN CELE.

OS THERMOMONOSPORA FUSCA.

OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;

OC ACTINOMYCETALES; STREPTOSPORANGINEAE; NOCARDIOPSACEAE; THERMOBIFIDA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-YX;

RX MEDLINE; 91258320.

RA LAO G., GHANGAS G.S., JUNG E.D., WILSON D.B.;

RT "DNA sequences of three beta-1,4-endoglucanase genes from

RJ *Thermomonospora fusca*.";

RL R. BACTERIOL. 173:3397-3407(1991).

CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -!- PATHWAY: CELLULOSE DEGRADATION.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L01577; G154694; -.
CC PIR; C42360; C42360.
CC PROSITE; PS00659; GLYCOSYL_HVDROL_F5; 1.
CC PRAM; PF00150; cellulase; 1.
CC PFAM; PF00553; CBD_1; 1.
CC HSP; P07986; IEXH.
CC KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
CC FT SIGNAL 1 36
CC ET CHAIN 37 466 ENDOGLUCANASE E-5.
CC FT ACT_SITE 299 299 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
CC SQ SEQUENCE 466 AA; 49760 MW; 114A0A0E CRC32;

CC Query Match 23.5%; Score 805; DB 1; Length 466;
CC Best Local Similarity 39.9%; Pred. No. 1.96e-132;
CC Matches 122; Conservative 81; Mismatches 88; Indels 15; Gaps 11;

Db 165 VERYGKVGQVGTQLCDHGPNVLGRGMSTHGIGWDFCHLTDSLDALEYDWKADIIRLSM 224
Cc | | : : : : : | | | | | | | | : : : : : | | : : | |
Qy 32 VEHQGLSINGELNVEGGVQLKGSSHLGWYGFVNYESKMWLRDDWGITVFRAAM 91

Db 225 YIQEDGYETNPRTGTRIDLDMATARGLYIVVDWHILTPGDPH-YNLDRAKTFFFAEA 283
Cc | | : : : : : | | | | | | | | : : | | | | : :
Qy 92 YTSSGGYIDDP-SVKERKVETVEAADLGIYVIDWHILSDNDPNIYK-EKADFPDEMS 149

Db 284 QRHASKTNLYEIANEENG--VSWAS-IKSYAEVPIVRQRDPDSVIIVGTGWSLGV 340
Cc : : : : : | | : : : : : | | : : | | | | | : : : : : | |
Qy 150 ELXGDYPNVIIYEIANEPNGSDVTWDNQIKPYAEVPIVRDNDPNNIIVGTGTWSQ-DV 208

Db 341 SEGSGPAEIAANPVNASINWAFHYAAASHRDNYLNALREASEL-FPVFVTEFTETYTG 399
Cc : : : : : | | | | | | | : : : : : | | : : | | : : | |
Qy 209 HHARD-NQLA-DP---NMVYAEHFYAGTQGRLRDQVDYALDQGAIFVSEWGTSA TG 262

Db 400 DGANDFOADRYIDLMAERKIGTKWNYSDDFRSGAVFGPTCASGPGWSGSLKASGOW 459
Cc | | : : : : : | | : : : : : | | : : | | : : : : : | |
Qy 263 DGGVFLEDAQWIIDFMFDERNLSWANWSLTHKDESSAALMPGANPTGG-WTEAELSPTGTF 321

Db 460 VRSLQ 465
Cc | | : :
Qy 322 VREKIR 327

RESULT 14
ID GUN_BACSI STANDARD; PRT; 800 AA.
AC P06564;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (ALKALINE CELLULOSE).
OS BACILLUS SP. (STRAIN 1139).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
RN BACILLUS.
RP [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87085443.
RA FUKUMORI F., KUDO T., NARAHASHI Y., HORIKOSHI K.;
RT "Molecular cloning and nucleotide sequence of the alkaline cellulase
RT gene from the alkalophilic Bacillus sp. strain 1139.";

RL J. GEN. MICROBIOL. 132:2329-2335(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- ALKALOPHILIC BACILLUS SP., STRAIN 1139, PRODUCES ONE ALKALINE
CC CELLULASE (PH OPTIMUM 9.0). THIS STRAIN IS NOT A TRUE CELLULOXYLIC
CC MICRO-ORGANISM BECAUSE THE ENZYME IS UNABLE TO HYDROLYSE NATIVE
CC CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M15743; GI42667; -
CC EMBL: D00066; D1000485; -
CC FIR: A29003; A29003.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC PFAM: PF00150; cellulase; 1.
CC KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
CC FT SIGNAL 1 30 POTENTIAL.
CC FT CHAIN 31 800 ALKALINE CELLULASE.
CC FT ACT_SITE 190 190 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 305 305 NUCLEOPHILE (BY SIMILARITY).
CC SQ SEQUENCE 800 AA; 88602 MW; 547179BA CRC32;

Query Match 21.6%; Score 739; DB 1; Length 800;
Best Local Similarity 40.5%; Pred. No. 6.52e-119;
Matches 126; Conservative 76; Mismatches 79; Indels 30; Gaps 15;

Db 68 LVDOHGKIQQLGMSHGLQWFFPEILNAYKALANDWESMIRLAWYGENGYASNPGL 127
Qy 45 LVNEGQVQLGMSHGLQWGFVNYESMKWLRDNGITVFRAMITSSGGYIDDPG- 103
Db 128 IKSRIKIGDLAIENDMYIVDWHVHAPGDPDPVY-AGAEDFFRDIATLPPNPHIYE 186
Qy 104 VKEKVKETVEAIDLGIYVIDWHILSDNDP-N-IYKEAKDFDEMSELYGDPNVIYE 161
Db 187 LANEPSSNNGGAGIPNNEGNNAVKEYADPIVEMLRDSNADNDNIIVGSPNWSQPD 246
Qy 162 IANEP---NG-SDVTW--D--NQIKPYAEVIVIRDN-DPN-NIVIGTGTWSQDVH 210
Db 247 ADNPIDDDHTHTYVHYFGSHAATESYPPETPNSERGVNMNTRYALENGVAVFATEW 306
Qy 211 AADNLQADPNVYAFHYAGTHG-----Q--NL-RDQV--D--YALDQGAIFVSEW 255
Db 307 GTSQANGDGGPVFDEADVWIEFLNENNISWANSWLNKNEVSGAFTPFELGKSNATSLDP 366
Qy 256 GTSATGCGGVFLDEAQVWIDFMDERNLSWANSWLNTHKDESSAALMPGANPTGWTAE 315
Db 367 GPDQVWVPEEL 377
Qy 316 SPGSGTFVREKI 326

RESULT 15
ID GUNA_STRLI STANDARD; PRT; 459 AA.
AC P27035;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE CELA PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).
GN CELA.
OS STREPTOMYCES LIVIDANS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-43.

RC STRAIN=66;
RX MEDLINE: 92246492.
RA THERBERGE M., LACAZE P., SHARECK F., MOROSOLI R., KLUEPFEL D.;
RT "Purification and characterization of an endoglucanase from
RT Streptomyces lividans 66 and DNA sequence of the gene.";
RL APPL. ENVIRON. MICROBIOL. 58:815-820(1992).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- FTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC SITE FOR PROTEOLYSIS.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M82807; -; NOT_ANNOTATED_CDS.
CC DR PROSITE: PS00561; CBD_BACTERIAL; 1.
CC DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC DR PFAM: PF00150; cellulase; 1.
CC DR PFAM: PF00553; CBD_1; 1.
CC DR HSP: P07986; 1EXH;
CC KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
CC FT SIGNAL 1 27
CC FT CHAIN 28 459 ENDOGLUCANASE CELA.
CC FT DOMAIN 136 135 CELLULOSE-BINDING.
CC FT DOMAIN 147 147 LINKER ("HINGE") (PRO-THR BOX).
CC FT DOMAIN 148 357 CATALYTIC.
CC FT ACT_SITE 286 286 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
CC FT DISULFID 31 131 BY SIMILARITY.
CC SQ SEQUENCE 459 AA; 48663 MW; 4035BEC4 CRC32;

Query Match 21.3%; Score 727; DB 1; Length 459;
Best Local Similarity 39.7%; Pred. No. 1.84e-116;
Matches 124; Conservative 71; Mismatches 98; Indels 19; Gaps 15;

Db 156 NGQLHVCVHLCNOYDRPIQLRGMSTHGIOWFGPCYGD-ASLDRLAODKSDLLRVANV 214
Qy 35 HGQUSISNGELVNERGEQVQLKGMSSHGLQWYQ-FVNYESMKWLRDNGITVFRAMIT 93
Db 215 QEDGYETDPAGFTSRVNGLVDMADRGMYAVIDFHTLTPGDPN-YNLDRARTFSSVAAR 273
Qy 94 SSGGYIDDPV-KKKVETVEAIDLGIYVIDWHILSDNDPNYK-EAKDFDEMSEL 151
Db 274 -NDKKNVIYEIANEPNG--VSWTA-VKSYAEQVIVPRAADPDVAVVIGTRGWSGLGVD 329
Qy 152 YGDPNVIYEIANEPNGSDVTWQIKPYAEVIVIRDNPNVIVIGTGTWSQ-DVHH 210
Db 330 GANESEVNNPVPVATNIMYAFHYAASHKDYRAVRPAATR-LPLFVSEFGTVSATAWS 388
Qy 211 AA-DNQLAD-P-----NVMYAFHYAGTHGQNLRDQVYALDQGAIFAIFSENGTSAATGDG 264
Db 389 -VDRSSSVAVLDDLQKLSYANTYSDADSGSAFREGTCGTGTDYSSSGVLTSSGALVK 447
Qy 265 GVFLDEAQVWIDFMDERNLSWANSWLNTHKDESSAALMPGANPTGWTAE-LSPSGTFVR 323
Db 448 SRISTTDDFFTS 459
Qy 324 EKRESASIPPS 335

Search completed: Fri Jun 4 09:56:15 1999
Job time : 25 secs.

WQESREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 4 09:56:32 1999; MasPar time 29.15 Seconds
874.368 Million cell updates/sec
Tabular output not generated.

Title: >US-08-945-574-1
Description: (1-467) from US08945574.pep
Perfect Score: 3419
Sequence: 1 MKKITTIFAVLLMTLALFSI.....KWTQNPDPGDPYGPWEPLN 467

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprenb19

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organella
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 49.047; Variance 101.519; scale 0.483

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	2781	82.3	400	2	085465	ENDOGLUCANASE-CELL	0.00e+00
2	2578	75.4	389	2	059232	ENDO-BETA-1,4-GLUCANAS	0.00e+00
3	1649	48.2	499	2	045532	CELLULASE.	1.33e-297
4	1643	48.1	499	2	052731	ENDO-B-1,4-GLUCANASE (2.23e-296
5	1637	47.9	486	2	045430	ENDOGLUCANASE PRECURSO	3.72e-295
6	1622	47.4	501	2	083012	CELLULASE.	4.24e-292
7	1594	46.6	387	2	031029	BETA(1,4)-GLUCAN GLUCA	2.14e-286
8	942	27.6	570	2	059665	ENDO-1,4-BETA-GLUCANAS	9.07e-155
9	895	26.2	749	2	059154	ENDOGLUCANASE (EC 3.2.	2.07e-145
10	850	24.9	557	2	P94622	ENDO-1,4-BETA-GLUCANAS	1.78e-136
11	853	24.9	930	2	059290	BETA-1,4-ENDOGLUCANASE	4.52e-137
12	840	24.6	481	2	066084	ENDOGLUCANASE.	1.71e-134
13	830	24.3	635	2	066065	CMC-XYLANASE (FRAGMENT	1.63e-132
14	749	21.9	783	2	045554	ENDOGLUCANASE PRECURSO	1.58e-116
15	737	21.6	476	5	018453	BETA-1,4-ENDOGLUCANASE	3.61e-114
16	739	21.6	821	2	059241	ENDO-1,4-BETA-GLUCANAS	1.46e-114
17	707	20.7	319	5	077449	BETA-1,4-ENDOGLUCANASE	2.78e-108
18	701	20.5	391	5	077094	BETA-1,4-ENDOGLUCANASE	4.16e-107
19	701	20.5	392	5	044078	BETA-1,4-ENDOGLUCANASE	4.16e-107
20	697	20.4	319	5	061595	BETA-1,4-ENDOGLUCANASE	2.52e-106

21	699	20.4	470	5	016028	BETA-1,4-ENDOGLUCANASE	1.02e-106
22	691	20.2	319	5	018454	BETA-1,4-ENDOGLUCANASE	3.77e-105
23	675	19.7	494	2	086099	CELLULASE PRECURSOR (E	5.05e-102
24	648	19.0	148	2	Q03882	ENDOGLUCANASE PRECURSO	5.92e-97
25	563	16.5	621	2	007653	CELLULASE B.	2.64e-80
26	507	14.8	910	3	P87211	ENDOGLUCANASE PRECURSO	1.39e-69
27	202	5.9	363	2	007652	CELLULASE A.	1.37e-14
28	193	5.6	516	2	086185	MANNANASE.	3.83e-13
29	176	5.1	357	2	Q59441	CELODEXTRINASE PRECUR	1.80e-10
30	170	5.0	865	2	Q43919	CHITINASE A PRECURSOR	1.50e-09
31	169	4.9	287	2	Q59144	CHITINASE PRECURSOR (E	2.13e-09
32	163	4.8	507	2	050506	PROBABLE HYDROLASE.	1.73e-08
33	160	4.7	252	2	086614	PUTATIVE SECRETED CHIT	4.87e-08
34	160	4.7	474	2	Q59142	CHITINASE PRECURSOR (E	4.87e-08
35	161	4.7	542	2	Q59145	CHITINASE II PRECURSOR	3.45e-08
36	157	4.6	538	2	Q59143	CHITINASE PRECURSOR (E	1.36e-07
37	154	4.5	729	2	P96168	CHITINASE A.	3.77e-07
38	153	4.5	831	2	050076	CHITINASE B (EC 3.2.1.	5.28e-07
39	151	4.4	535	2	Q59141	CHITINASE PRECURSOR (E	1.04e-06
40	151	4.4	665	2	Q48373	CHITINASE PRECURSOR.	1.04e-06
41	141	4.1	288	2	Q52246	ORFC.	2.83e-05
42	139	4.1	294	2	Q50152	CHITINASE C.	5.41e-05
43	138	4.0	268	2	Q54942	ORF IOTA.	7.47e-05
44	137	4.0	695	2	Q52863	EXOCHITINASE.	1.03e-04
45	134	3.9	430	2	Q50590	CHITINASE PRECURSOR (E	2.69e-04

ALIGNMENTS

RESULT 1
ID 085465 PRELIMINARY; PRT; 400 AA.
AC 085465;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ALKALINE CELLULOSE CEL5A (EC 3.2.1.4).
GN CEL5A.
OS BACILLUS AGARADHAERENS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 8721;
RA BJORNVAD M.E.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF067428; G3193320; -;
KW HYDROLASE; GLYCOSIDASE.
SQ SEQUENCE 400 AA; 44702 MW; 85AEALF6 CRC32;

Query Match 81.3%; Score 2781; DB 2; Length 400;
Res. Local Similarity 95.6%; Pred. No. 0.00e+00;
Matches 387; Conservative 8; Mismatches 5; Indels 5; Gaps 1;

Db	1	MKKITTIFVLLMTVALFSIGNTTAAANDSVVEHGQLSISNGELVNERGEQVQLKGMS	60
Qy	1		60
Db	61	HGLWYGFVYVESMKVLRDDWGINVFRAMYTSSGGYIDDPVSKEKVEAVEAIDLDI	120
Qy	61		120
Db	121	YVIDWHLSDNDPNVKEEAKDFDEMSLYGDPVNIYEIANEPNGSDVTWGNQIKPY	180
Qy	121		180
Db	181	AEVPIIRNDPNIIIVGTGTSQDVHHAADNQLADPNVYAFHFYAGTHGQNLQOV	240
Qy	181		240
Db	241	DYALDQGAIFVSEWGTSAATGDGVFLDEAQVWIDFMDERNLSWANNSLTHKDESSAAL	300
Qy	241		300

RA YOON K.-H., JUNG K.H.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
DR EMBL; AF045482; G2854064; -.
KW HYDROLASE; GLYCOSIDASE.
SQ SEQUENCE 499 AA; 55193 MW; 114187B8 CRC32;
Query Match 48.1%; Score 1643; DB 2: Length 499;

Db	61	GISSHGLQWGEVYVNNKDSIKWLRRDDGWIIVFRAAMYATDGGYIDNPSSVKNVKEAVEAK	120
QY	57	GNSSHGLQWQGVNYESMKWLRDDGWIIVFRAAMYSGGYIDDPVSKEVKETVEAAI	116
Db	121	ELGIYIIDWHIINDGPNONKEKAEFFKEMSSLYGNTPNVYEIETANEPNG-DVNNKRD	179
QY	117	DLGIYIIDWHIILSDNPNIYKEAKDFDEKSELYGDPNVYIETANEPNGSDVTWQ	176
Db	180	IKPYAEVSVIRKRPDNIIVGTGTWSQDVNDAAADQDKDANVMYALHFYAGTHGQFL	239
QY	177	IKPYAEVPIVRDNDPNVIIVGTGTWSQDVHHAADNLADPNVMYAFHFYAGTHGQNL	236
Db	240	RDKANVALSKGAPFYETWGTSDASNGGVFLDOSREWLKILDSKTIISVNNWNLSDKQES	299
QY	237	RQVDYDALDGAALFYSEWGTSAATGGGVFLDEAQWIDFMDERNLSWANNSLTHKDES	296
Db	300	SSALKPGASKTGGWRULSDLSASGTFVRENILGTKDTKDIPTPSKDKTQEBNG	353
QY	297	SAALMGFANPTGGWTAELISPSGTFYREKIRISASITPPSDP-TPPSDGPEPPG	349

DE CELLULOSE.
OS BACILLUS SP.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
CC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5H;
RA KHANONGNUCH C., OOI T., KINOSHITA S.;
RT "Cloning and nucleotide sequence of beta-mannanase and cellulase gene
from *Bacillus* sp. 5H";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB016164; D1032673;
SQ SEQUENCE 501 AA; 55422 MW; 5D38F2BD CRC32;

D6 300 SSALKPGASTKGWPLTDLTASCTFYRENIFGNKDSTKERPETPAQDNPAQENG 353
I::IIII: IIII :::IIIIIII: I : II: I :
QY 297 SAALMPCANTGCGTAEALSPTSCTFVREKI-RESASIPSDPTPPSDPGEPPDG 349

RESULT 7
ID O31029 PRELIMINARY; PRT; 387 AA.
AC O31029;

RESULT 8

ID	Q59665	PRELIMINARY;	PRT;	570 AA.
AC	Q59665;			
DT	01-NOV-1996	(TREMBLREL. 01, CREATED)		
DT	01-NOV-1996	(TREMBLREL. 01, LAST SEQUENCE UPDATE)		
DE	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	ENDO-1,4-BETA-GLUCANASE	(EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)		
DE	(CARBOXYMETHYL CELLULOSE).			
GN	CELE.			
OS	PSEUDOMONAS FLUORESCENS.			
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;			
OC	PSEUDOMONAS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 95366948.			
RA	HALL J., BLACK G.W., FERREIRA L.M.A., MILLWARD-SADLER S.J.,			
RA	ALI B.R.S., HAZLEWOOD G.P., GILBERT H.J.;			
RT	"The non-catalytic cellulose-binding domain of a novel cellulase from			
RT	Pseudomonas fluorescens subsp. cellulosa is important for the			
RT	efficient hydrolysis of Avicel.";			
RL	BIOTECHEM. J. 309:749-756(1995).			
CC	-I- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC			
CC	LINKAGES IN CELLULOSE.			
DR	EMBL; X86798; G808574; -.			

RESULT	9	
ID	Q59154	PRELIMINARY; PRT; 749 AA.
AC	Q59154;	
DT	01-NOV-1996	(TREMBLREL. 01, CREATED)
DT	01-NOV-1996	(TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE	ENDOGLUCANASE (EC 3.2.1.4)	(CELLULOSE) (ENDO-1,4-BETA-GLUCANASE)
DE	(CARBOXYMETHYL CELLULOSE).	
GN	CELD.	
OS	ANAEROCELLUM THERMOPHILUM.	
OS	BACTERIA; FRMNICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;	
OC	ANAEROCELLUM GROUP; ANAEROCELLUM.	
EN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-2-1320;	
RA	ZVERLOV V., ASCHERL G., VELIKODVORSKAYA G., BRONNEMEIER K.;	
RL	SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.	
CC	-!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC	
CC	LINKAGES IN CELLULOSE.	
DR	EMBL; 277855; E257608; --	
DR	PFAM; PF00150; cellulase; 1.	
DR	PFAM; PF00395; SLH; 3.	
KW	HYDROLASE; GLYCOSIDASE.	
SQ	SEQUENCE 749 AA; 85066 MW; C558A863 CRC32;	

Query Match	26.2%	Score 895;	DB 2;	Length 749;
Best Local Similarity	46.5%	Pred. No. 2.07e-145;		
Matches	125;	Conservative 66;	Mismatches 65;	Indels 13; Gaps 9;

Db	68	LADQKGFIILRGMTSHGLQWYGDIIINKNFAKLSKDWCNVIRLAMYGEGGYASNPSI	127
		: : : : : : : : : : : : : : : : : : : : : : : :	
Qy	45	LVNERGEQVQLKGSSHGLQWYGOFVNYESKMWLRDDWGITVRAAMYTSSGGYIDDPV	104
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	128	KEKVEIGIKLAIENDMYIVDVHVLNPGDPNAYIKG-KADFFKEIATGFPNDY-HIIVE	185
		: : : : : : : : : : : : : : : : : : : : : : : :	
Qy	105	KEKVKETVEAAILDGIIVIIDWHILSDNDPN--IYKEEAKDFDENSELY-GDYPNVIYE	161
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	186	LCNPEPNPEGCVENSLDGWKVKYAYAQPIITKMLRSLGNOHIIIVGSPNNWSQRDFAIQDP	245
		: : : : : : : : : : : : : : : : : : : : : : : :	

QY 162 IANEPGSD--V--TWD--NOIKPYAEVPIVRDNDPNNIVIGTGTWSDVHHAADNQ 215
 Db 246 INDKNWMYVHFVSGTHKVGIVFENKKNAPFVPSVSGTSLASGDGGPYLDEADK 305
 QY 216 LADPNMYAFHFYAGTH--GQN--LRDQVDYALDQGAIFVSEWGTSAATGDGGVFLDEAQQ 273
 Db 306 WLEYLANSYISWNWSLSNKSNTSAFVP 334
 QY 274 WIDFMDERNLSWANWSLTHKDESSAALMP 302

RESULT 10
 ID P94622 PRELIMINARY; PRT; 557 AA.
 AC P94622;
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE ENDO-1,4-BETA GLUCANASE ENG (EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)
 DE (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
 GN ENGF.
 OS CLOSTRIDIUM CELLULOVORANS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 OC CLOSTRIDIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97136706.
 RA SHEWITA S.A., ICHI-ISHI A., PARK J.S., LIU C., MALBURG L.M.,
 RA DOI R.H.;
 RT "Characterization of engF, a gene for a non-cellulosomal Clostridium
 cellulovorans endoglucanase.";
 RL GENE 182:163-167(1996).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 DR EMBL; U37056; G1778709; -.
 DR PFAM; PF00150; cellulase; 1.
 KW HYDROLASE; GLYCOSIDASE.
 SQ SEQUENCE 557 AA; F4D7C185 CRC32;

Query Match 24.9%; Score 850; DB 2; Length 557;
 Best Local Similarity 41.3%; Pred. No. 1.78e-136;
 Matches 133; Conservative 76; Mismatches 91; Indels 22; Gaps 13;

Db 63 LCDKDGNIQLRGMSTHGLQWFGVGNVNNFAALSDNWSNVIRLAMYVABGGYATNP 122
 QY 45 LVNERGEQVQLKGMSSHGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGGYIDDP 104
 Db 123 KQTVINGINVAINDMTVIVDWHMNPEDPNASVYSG-AQSFNDISTLYPNNKNIYEL 181
 QY 105 KEVKETVEAADILGIYVVDHILSDNDPN--IYKEAKDFFDEMSELYGDPNVIYEI 162
 Db 182 CNEPENGCGVTNDATGWAQVKSATPIVOLLRNKGNENLIIVGNPFWQSRPDLAADNPI 241
 QY 163 ANEPGSD--VTWDN----QIKPYAEVPIVRDNDPNNIVIGTGTWSDVHHAADNQ 216
 Db 242 NDSNTYVHFVSGTPISTVTDNRNAMSNVRYALNHAAGVAFATWGTSLATGTGYP 301
 QY 217 ADPNVMYAFHFYAGTH--GQ-NL-RDQ-V-D--YALDQGAIFVSEWGTSAATGDGGVFL 268
 Db 302 AKADAWLDFLNGNNSISWNSISNKSNAALNSLTSLDPSGDKLWADNELLTSGQVYRA 361
 QY 269 DEAQVWIDFMDERNLSWANWSLTHKDESSAAL--MPGANPTGG--WTEAELSPSGTFVRE 324
 Db 362 RIKGAYATPVPVT-NQPTAP 382
 QY 325 KIRASASIPSPDPTPPSDGPGE 346

RESULT 11
 ID Q59290 PRELIMINARY; PRT; 930 AA.
 AC Q59290;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)

DE BETA-1,4-ENDOGLUCANASE (EC 3.2.1.4).
 GN CELA.
 OS CLOSTRIDIUM JOSUI.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 OC CLOSTRIDIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA FUJINO T., FUJINO E., KARITA S., OHMIYA K.;
 RT "Revised sequence of cea gene encoding endoglucanase (EG)-1 from
 Clostridium josui.";
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; D85526; D1013510; -.
 DR PFAM; PF00150; cellulase; 1.
 DR PFAM; PF00395; SLH; 1.
 KW HYDROLASE; GLYCOSIDASE.
 SQ SEQUENCE 930 AA; 101726 MW; B1C65689 CRC32;

Query Match 24.9%; Score 853; DB 2; Length 930;
 Best Local Similarity 44.2%; Pred. No. 4.52e-137;
 Matches 122; Conservative 71; Mismatches 37; Indels 20; Gaps 14;

Db 62 LCDKDGNIQLRGMSTHGLQWFGVGNVNNFAALSKDGSNVIRLAMYVABGGYSKDPEI 121
 QY 45 LVNERGEQVQLKGMSSHGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGGYIDDP 104
 Db 122 IKKRVIDGIDLAINDMTVIVDWHVLTGPDNPADYK--AMDFFKEISOKYNNPHIYE 180
 QY 105 -KEVKETVEAADILGIYVVDHILSDNDPN--IYKEAKDFFDEMSELYGDPNVIYE 161
 Db 181 LANEPSPNDPGVTNDADGAWK--VKSAYEPIIKLRDSNKNLIIVGSPNWSQRPDLAAEN 239
 QY 162 IANEP--N--G-S-DVT--WDNQIKPYAEVPIVRDNDPNNIVIGTGTWSDVHHAADN 214
 Db 240 PINDNNTAYSPHFYSGTHKSTSTSDTRGNIMSNARYALEHGVAVFCSEWGTSEAGNNCP 299
 QY 215 QLADPNVMYAFHFYAGTH--G-QNL-RDQV--D--YALDQGAIFVSEWGTSAATGDGGV 266
 Db 300 YLKEADWLEFLNANNISWINNLSLTKNETSGSFIP 335
 QY 267 FLDEAQVWIDFMDERNLSWANWSLTHKDESSAALMP 302

RESULT 12
 ID O66064 PRELIMINARY; PRT; 481 AA.
 AC O66064;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE ENDOGLUCANASE.
 OS ACTINOMYCES SP. 40.
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
 OC ACTINOMYCETALES; ACTINOMYCINAEAE; ACTINOMYCETACEAE; ACTINOMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-40;
 RA CHO K.K.;
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U94825; G2980982; -.
 SQ SEQUENCE 481 AA; 50751 MW; 6DE4F93F CRC32;

Query Match 24.6%; Score 840; DB 2; Length 481;
 Best Local Similarity 37.6%; Pred. No. 1.71e-134;
 Matches 120; Conservative 83; Mismatches 107; Indels 9; Gaps 7;

Db 44 GLSAPTILDEHGMPPQLRGASTHGLQWFGVGNVNNFAALSDNWSNVIRLAMYVABGGY 103
 QY 39 SISNGELVNERGEQVQLKGMSSHGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGGY 98
 Db 104 LOGSQAQMDKTIQCGVQAATDLGVTIIDHVL--NYPNGDATQAESFFKSYAAKYSVG 162
 QY 99 IDDPSPV-KEK-VKETVEAADILGIYVVDHILSDNDPNVIYKEAKDFFDEMSELYGDP 156
 Db 163 NVIEVCNEPTGTPWYDGSNDIYSYCTMAKAIKADAGSDAILCGTNTWSDIDAVAGK 222

```

157 NVIYEIANEPNGSD--VTWQNKIPYAEVPIVRDNDPNNIVIGTGTWSQDVHHAADN 214
223 PLSADGDNMTMYVLFHYAANTHKKDLRAKLTALNAGTPVFVSEFGCLDASNGGIDQDSA 282
215 QL-ADP--NVMYAFHFYAGTHGQNLROVDYALDQGAIFVSEWGTSAATGGGVFLDEA 271
283 NAWMTLLAHNNISYAANALSKAETAFAFKSPKTSKWTGDDTTPSAIMLVNTSRKLAD 342
272 QVWIDFMDERNLWANNLSLTHKDESSALMPGANPTGGWTEAEILSPSGTFFVREKIREAS 331
343 FADHAASGTSSGSKASSG 361
332 IP-PSDTPPSDGPDPG 349

RESULT 13
ID O66065 PRELIMINARY; PRT; 635 AA.
AC O66065;
DT 01-AUG-1998 (TREMBLREL, 07, CREATED)
DT 01-AUG-1998 (TREMBLREL, 07, LAST SEQUENCE UPDATE)
DE CMC-XILANASE (FRAGMENT).
OS FIBROBACTER SUCCINOGENES S85.
OC BACTERIA; FIBROBACTER GROUP; FIBROBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S85;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U94826; G2980984; -
KW XILAN DEGRADATION; HYDROLASE; GLYCOSIDASE.
FT NON_TER 635
SQ SEQUENCE 635 AA; 66943 MW; 2510AB57 CRC32;

Query Match 24.3%; Score 830; DB 2; Length 635;
Best Local Similarity 38.0%; Pred. No. 1.63e-132; Indels 10; Gaps 8;
Matches 127; Conservative 80; Mismatches 117;

Db 78 GFSAPTVDLHGQFQLRGASTHGVQWFFQYINRDFAQSLRDEWGINMVRILALYPREGGY 137
QY 39 SINGELVNERGEQVLKGMSSHGLQYGFVNYESMKWLDDGIGITVFRAMYTSSGGY 98
Db 138 LQSQAKMDKIEAVNAELGMVYLDHVL-NYNPBGADAAEFTFYATKYKWLK 196
QY 99 IDDPVSK-E-KVKTVEAAIDGLIYIWDHILSDNDPNYKEAKDFDEMSELYGDP 156
Db 197 NVLYEIDNEPTSTSWYDGSNDLTYSKRTKATKATGNQSNQSVICGTWTWSQDVDAVAK 256
QY 157 NVIYEIANEPNGSD--VTWQNKIPYAEVPIVRDNDPNNIVIGTGTWSQDVHHAADN 214
Db 257 PLSADGIGNVAYTLHFYAGTHYDNKIKRLTALAAGTPVFVSEFGITDASGWGGIDIANA 316
215 QL-ADP--NVMYAFHFYAGTHGQNLROVDYALDQGAIFVSEWGTSAATGGGVFLDEA 271
317 NDWMTLLTRNNISYAANSLCNKGEGAFLEKSTSKTSKWTGSEISTGIMLVNTSRRTQA 376
QY 272 QVWIDFMDERNLWANNLSLTHKDESSALMPGANPTGGWTEAEILSPSGT-FVREKIRESA 330
Db 377 MVDVVSQSGSGTGET-PTDTPDTPDTPDPT 409
QY 331 SIPPSPDTPPSDGPDPGPDPTPPSDPGEYPA 364

RESULT 14
ID Q45554 PRELIMINARY; PRT; 783 AA.
AC Q45554;
DT 01-NOV-1996 (TREMBLREL, 01, CREATED)
DT 01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
OS (ALKALINE CELLULOSE).
OC BACILLUS SP.

```

```

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22-28;
RA MIYATAKE M., IMADA K.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A
(FAMILY 5 OF GLYCOSYL HYDROLASES).
CC EMBL; D85236; D1013426; -
DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
DR PFAM; PF00150; cellulase; 1.
KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 783 ALKALINE CELLULOSE.
FT ACT_SITE 190 190 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 305 305 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 783 AA; 86964 MW; C95F61BC CRC32;

Query Match 21.9%; Score 749; DB 2; Length 783;
Best Local Similarity 42.5%; Pred. No. 1.58e-116;
Matches 122; Conservative 75; Mismatches 60; Indels 30; Gaps 14;

Db 68 LYDQGEKIQLRGMSHGLQWFPFELNDNAYKALSDNDWNSMIRLAMYGVNGYATNPDL 127
QY 45 LVNERGEQVLKGMSSHGLQYGFVNYESMKWLDDGIGITVFRAMYTSSGGYDDPS- 103
Db 128 IKORVIDGIELAIENDMVIYVDWHVHAPGDPDPVY-AGAEDFFRDIAALYPNPHIYE 186
QY 104 VKEKVEIVEAAILGLIYIWDHILSDNDP-N-IYKEAKDFDEMSELGDPNVIYE 161
Db 187 LANEPSSNNNGAGIPNNEEGKAVKEYADPIVEMLRDSGNADNIIIVSPNWSQRPDL 246
QY 162 IANEP---NG-SDVTWQDQ---IKPYAEVPIVRDNDP-NIVIVGTGTWSQDVH 210
Db 247 AADNPINDHHTMYVHFVTGSHAATESYPPETPNSEGNVMSNTRYALENGVAVFATEW 306
QY 211 AADNLQADPNYAFHFYAGTHG-----Q--NL-RDQV--D--YALDQGAALFVSEW 255
Db 307 GTSQANGDGGYFDEADVWIEFLNENNISWANSLTNKNEVSGATFP 353
QY 256 GTSATGDDGVFLDEAQVWIDFMDERNLWANNLSLTHKDESSAALMP 302

RESULT 15
ID O18453 PRELIMINARY; PRT; 476 AA.
AC O18453;
DT 01-JAN-1998 (TREMBLREL, 05, CREATED)
DT 01-JAN-1998 (TREMBLREL, 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL, 08, LAST ANNOTATION UPDATE)
DE BETA-1,4-ENDOGLUCANASE-1 PRECURSOR (EC 3.2.1.4) (CELLULOSE)
DE (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
GN HG-ENG-1.
OS HETERODERA GLYCINES.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; DIPLOGASTERIA; TYLENCHIDA;
OC TYLENCHINA; TYLENCHOIDEA; HETERODERIDAE; HETERODERINAE; HETERODERA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98226740.
RA SWANT G., STOKKERMANS J.P.W.G., YAN Y., DE BOER J.M., BAUM T.J.,
RA WANG X., HUSSEY R.S., GOMMERS F.J., HENRISSAT B., DAVIS E.L.,
RA HELDER J., SCHOTS A., BAKKER J.;
RT "Endogenous cellulases in animals: isolation of beta-1,
RT 4-endoglucanase genes from two species of plant-parasitic cyst
RT nematodes.";
RL PROC. NATL. ACAD. SCI. U.S.A. 95:4906-4911(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OF50;
RA YAN Y., SWANT G., STOKKERMANS J.W.P.G., WANG X., HUSSEY R.S.,
RA BAKKER J., HELDER J., SCHOTS A., DAVIS E.L.;

```


2
2
0
2
-
-
-
0
3
4
0
2
0
0

```

MPPerch_pp  protein - protein database search, using Smith-Waterman algorithm
Run on:      Fri Jun 4 09:58:07 1999; Maspar time 20.71 Seconds
              903.345 Million cell updates/sec
Tabular output not generated.

```

Scoring table: PAM 150
Gap 11

Post-processing: Minimum Match 0%
Listing first 45 summaries

```

data333:
    p1r0
      1:p1r1 2:p1r2 3:p1r3 4:p1r4

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query		Length	DB	ID	Description	Pred. No.	
	Match	Score					EC	
1	82.7	2828	409	2	B25156	cellulase (EC 3.2.1.4	0.00e+00	
2	82.0	2803	488	2	A25156	cellulase (EC 3.2.1.4	0.00e+00	
3	1664	48.7	505	2	S39962	endoglucanase - Erwin	4.29e-266	
4	1660	48.6	508	2	A26874	cellulase (EC 3.2.1.4	4.28e-266	
5	1649	48.2	499	2	A27198	cellulase (EC 3.2.1.4	2.26e-266	
6	1641	48.0	508	2	G69593	cellulase (EC 3.2.1.4	6.37e-266	
7	1637	47.9	486	2	I40548	bifunctional cellulase	3.39e-266	
8	1616	47.3	504	2	S54744	cellulase (EC 3.2.1.4	2.17e-255	
9	1598	46.7	499	2	JN0111	cellulase (EC 3.2.1.4	3.97e-253	
10	1491	43.6	448	2	A27631	cellulase (EC 3.2.1.4	9.24e-233	
11	942	27.6	570	2	S56332	cellulase (EC 3.2.1.4	2.65e-133	
12	884	25.9	429	2	S29044	endoglucanase A precu	4.93e-122	
13	850	24.9	557	2	JC5487	cellulase (EC 3.2.1.4	4.96e-122	
14	805	23.5	466	2	C42360	cellulase (EC 3.2.1.4	4.16e-111	
15	793	23.2	428	2	JC3767	cellulase (EC 3.2.1.4	5.33e-111	
16	749	21.9	783	2	JC5467	cellulase (EC 3.2.1.4	2.73e-108	
17	739	21.6	800	2	A29003	cellulase (EC 3.2.1.4	1.54e-108	
18	739	21.6	822	2	JT0611	cellulase (EC 3.2.1.4	1.54e-108	
19	690	20.2	357	2	PC4404	cellulase (EC 3.2.1.4	5.42e-93	
20	690	20.2	941	2	S29043	cellulase (EC 3.2.1.4	5.42e-93	
21	613	17.9	825	2	J50174	cellulase (EC 3.2.1.4	1.17e-79	
22	196	5.7	516	2	JE0134	mannan endo-1,4-beta-	8.75e-12	
23	186	5.4	584	2	J01229	cellulase (EC 3.2.1.4	2.26e-10	

ALIGNMENTS

ENTRY	B25156	#type complete
TITLE	cellulase (EC 3.2.1.4) 2 -	Bacillus sp.

```

DATE
15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change
20-Mar-1998

```

REFERENCE

#journal

#accessio

#res#

```
##exp
```

#descript

KEYWORDS

○ ○ ○ ○ ○

Best Local Websites

1 M

QY I ME

DT 6T HC

УДК 62-50

177

T T T Ty

—
H
O
H
2
2

11

```
Db 241 DYALDQGAATFVSEWGTSEATGDCGVFLDEAQWIDFMDERNLNSWANSLTHKDESSAAL 300
QY 241 DYALDQGAATFVSEWGTSEATGDCGVFLDEAQWIDFMDERNLNSWANSLTHKDESSAAL 300
Db 301 MPGASPTGCGWTEAEALSPSGTFVREKIRSATPPSDPTPPSDPDGCEPDGPDGPDPTPP 360
QY 301 MPGANPTGCGWTEAEALSPSGTFVREKIRSATPPSDPTPPSDPDGCEPDGPDGPDPTPP 356
Db 361 SDPGDYPAMDNTIYTDIEIVYHNGOLQAKWWTQNOEGDYPGPWEPLN 409
QY 357 SDPGEYPAWDSNQIYTNIEIVYHNGOLQAKWWTQNOEGDYPGPWEPLK 405

RESULT 2 A25156 #type complete
ENTRY cellulase (EC 3.2.1.4) 1 - Bacillus sp.
TITLE endo-1,4-beta-glucanase
ALTERNATE_NAMES #formal_name Bacillus sp.
ORGANISM 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change
DATE 20-Mar-1998
ACCESSIONS A25156
REFERENCE A91825
#authors Fukumori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
#journal J. Bacteriol. (1986) 168:479-485
#title Nucleotide sequences of two cellulase genes from alkalophilic
Bacillus sp. strain N-4 and their strong homology.
#cross-references MUID:87056924
#accession A25156
#molecule_type DNA
#residues 1-488 #label FUK
#status preliminary
#cross-references GB:M14781; GB:X53449; NID:gl42659; PID:gl42660
#experimental_source strain N-4, plasmid pNK1

FUNCTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1,3-linkages
#pathway cellulose degradation
KEYWORDS glycosidase; hydrolase; polysaccharide degradation
SUMMARY #length 488 #molecular-weight 54264 #checksum 2043

Query Match 82.0%; Score 2803; DB 2; Length 488;
Best Local Similarity 84.0%; Pred. No. 0.00e+00;
Matches 414; Conservative 30; Mismatches 18; Indels 31; Gaps 9;

Db 1 MKKLTTFIVFTLAL-LF-VGNSTANNGSVVEQNGQLSIQGLVNEHGDVPVLKGMSS 58
QY 1 MKKLTTFIVFTLAL-LF-VGNSTANNGSVVEQNGQLSIQGLVNEHGDVPVLKGMSS 58
Db 59 HGLQWYGFVNYDSIKWLDDWGTIVFRAAMYTSGGYIEDPSVKEKVEAYEAAIDLGI 118
QY 61 HGLQWYGFVNYDSIKWLDDWGTIVFRAAMYTSGGYIEDPSVKEKVEAYEAAIDLGI 120
119 YVIDDWHLLSDNDPNIIYKEAKEFDEMSALYGDYPNVYIETANEPNGHNVRWDSHIRPY 178
QY 121 YVIDDWHLLSDNDPNIIYKEAKEFDEMSALYGDYPNVYIETANEPNGSDVTDWQIKRPY 180
179 AEEVPIVRANDPNNVIVTGVATWSODVHEAADNOLDPNVMYAFHFYAGTHGQQLRNQV 238
QY 181 AEEVPIVRANDPNNVIVTGVATWSODVHEAADNOLDPNVMYAFHFYAGTHGQQLRNQV 240
239 DYALSRGAATFVSEWGTSAATGDCGVFLDEAQWIDFMDERNLNSWANSLTHKDESSAAL 298
QY 241 DYALDQGAATFVSEWGTSAATGDCGVFLDEAQWIDFMDERNLNSWANSLTHKDESSAAL 300
299 MPGANPTGCGWTEAEALSPSGTFVREKIRSATPPSDPTPPSDPDGCEPDGPDPTPPSDPG 355
QY 301 MPGANPTGCGWTEAEALSPSGTFVREKIRSATPPSDPTPPSDPDGCEPDGPDPTPPSDPG 360
356 EYPADPNQIYTNIEIVYHNGOLQAKWWTQNOEGCANQYGPWEPLGDPAPSPSPSPSPSPS 415
QY 361 EYPADWSNQIYTNIEIVYHNGOLQAKWWTQNOEGP-DPYGPWEPL-K----SDP-D----S 410
```

```
Db 416 EPEPDGPDGPDGPDGPDGPDGPDGPDGPDGPDGPDGPDGPDGPDGPDGPDGPDGPDG 475
QY 411 -----GEPDPTPPSDGPEYPAWDSNQIYTNIEIVYHNGOLQAKWWTQNO 454
Db 476 EPGYDYPGPWEPLN 488
QY 455 EPGDYPGPWEPLN 467

RESULT 3 S39962 #type complete
ENTRY endoglucanase - Erwinia carotovora
TITLE #formal_name Erwinia carotovora
ORGANISM 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
DATE 09-Sep-1997
ACCESSIONS S39962
REFERENCE S39962
#authors Cooper, V.J.C.; Salmond, G.P.C.
#journal Mol. Gen. Genet. (1993) 241:341-350
#title Molecular analysis of the major cellulase (CelV) of Erwinia
carotovora: evidence for an evolutionary "mix-and-match" of
enzyme domains.
#accession S39962
#status preliminary
#molecule_type DNA
#residues 1-505 #label COO
#cross-references EMBL:X76000; NID:g434941; PID:g434942
SUMMARY #length 505 #molecular-weight 54899 #checksum 7815

Query Match 48.7%; Score 1664; DB 2; Length 505;
Best Local Similarity 67.7%; Pred. No. 4.29e-267;
Matches 216; Conservative 50; Mismatches 51; Indels 2; Gaps 2;

Db 35 VETHQQLSINGRLVDEQGRYQLRGISGHGLQWFGDYVYKMSKWLDRDWDGYNFVRAM 94
QY 32 VEEHQLSISNGELYNERGEQVQLKGMSSHGLQWGFVYVYKMSKWLDRDWDGIVFRAAM 91
Db 95 YTAADGYISNPLANKVKEAVAAASGLVYIIDWHILSDNDPNIIYKAQAKTFFAEMAGL 154
QY 92 YTSGGYIDDPVKEKVEAEADLGIYVIDWHILSDNDPNIIYKEEAKDFDEMSSEL 151
155 YCSSPNVYIETANEPNGG-VTWNGQIRPYALEVTDITRSKDPDNLIIYGTGTSQDIHDA 213
QY 152 YGDYPNVYIETANEPNGSDVTDWQIKPYAEVPIVRDNDPNNVIVTGTWSDQVHHA 211
214 ADNQLPDNPTMYALHFYAGTHGQFLRDIDYQAOSGAALFVSEWGTSDASGNGPFLPES 273
QY 212 ADNQLADPNMYAFHFYAGTHGQFLRDIDYQAOSGAALFVSEWGTSAATGGVFLDEA 271
274 QTWIDFLNNGYSWNWSLTDKSEASALAPGASKSGGWTEQNLSITSGKFVREQIRAGAN 333
QY 272 QWIDFMDERNLNSWANSLTHKDESSAALMPCANPTGGWTEAEALSPSGTFVREKIR 331
334 LGGGDTPTTPTPTPTNGNG 352
QY 332 IPPSD-PTPPSDGPDGPDG 349

RESULT 4 A26874 #type complete
ENTRY cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain
TITLE DLG)
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM #formal_name Bacillus subtilis
DATE 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change
10-Jul-1998
ACCESSIONS A26874; B26874
REFERENCE A26874
#authors Robson, L.M.; Chambliss, G.H.
#journal J. Bacteriol. (1987) 169:2017-2025
#title Endo-beta-1-4-glucanase gene of Bacillus subtilis DLG.
#cross-references MUID:87194581
```

```
##cross-references GB:M28332; NID:g142670; PID:g142671
##experimental_source strain IFO3034
FUNCTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1,3-linkages
#pathway cellulose degradation
KEYWORDS extracellular protein; glycosidase; hydrolase; polysaccharide
degradation
FEATURE
1-36 #domain signal sequence #status predicted #label SIG
SUMMARY #length 499 #molecular-weight 55075 #checksum 3211
Query Match 48.2%; Score 1649; DB 2; Length 499;
Best Local Similarity 62.7%; Pred. No. 2.26e-264;
Matches 222; Conservative 62; Mismatches 64; Indels 6; Gaps 6;
Db 1 MKRSISFTICLLITVTCGLQASPASAAAGTKTPAAKNQSLKIGTQLVNRDQKAVOLK 60
QY 1 MKKITTIFAV-LLMT-LALFSI-GNTTAADDYSV-VEEHGQLSISNGELVNERGEQVQLK 56
Db 61 GISSHGLQWYGFVYKESKWLKDDGKITVFRAMYTADGGYIDNPSVKNKVEAEAAK 120
QY 57 GMSHGLQWYGFVYKESKWLKDDGKITVFRAMYTSSGGYIDDPVSKVKETVEAAI 116
Db 121 ELGIYVIDWHILNDGNPNQKAKDFKEMSSLYGNTPNVIYEIANEPNG-DVNWKRD 179
QY 117 DLGIYVIDWHILNDGNPNYKKEAKDFDEMSELGYDNPVIYEIANEPNGSDVTWQNL 176
Db 180 IKPYAEVISVIRKNDPNIIIVGTGTSQVDAADQDLKANDVYALHYFAGTHGQSL 239
QY 177 IKPYAEVIVIRNDPNIIIVGTGTSQVDAADQDLKANDVYALHYFAGTHGQNL 236
Db 240 RDKANYALSKGAPFVTEWGTSDASNGGVFLDQSRWLNLYLDSKNISWYNWLSKQES 299
QY 237 RQVDYALDQGAIFVSEWGTSAATGGVFLDQSRWLNLYLDSKNISWYNWLSKQES 296
Db 300 SSALKPGASKTGWPLTDLTASGTFVRENIRGKDSKDPETPAQDNPTQKEG 353
QY 297 SAALMPGANPTGWTAEELSPSGTFVREKIRASIPSPDP-TPPSDPGEPDPG 349
RESULT 6
ENTRY #type complete
TITLE cellulase (EC 3.2.1.4) bg1C precursor - Bacillus subtilis
ALTERNATE_NAMES #formal_name Bacillus subtilis
ORGANISM #sequence_revision 05-Dec-1997 #text_change
DATE 05-Dec-1998
ACCESSIONS G69593; I40353; S24239; S49103; I39803
REFERENCE A69580
#authors
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Dueterhoeft, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galleron, N.; Gim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
```

```
##cross-references GB:M28332; NID:g142670; PID:g142671
##experimental_source strain IFO3034
FUNCTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1,3-linkages
#pathway cellulose degradation
KEYWORDS extracellular protein; glycosidase; hydrolase; polysaccharide
degradation
FEATURE
1-36 #domain signal sequence #status predicted #label SIG
SUMMARY #length 499 #molecular-weight 55075 #checksum 3211
Query Match 48.2%; Score 1649; DB 2; Length 499;
Best Local Similarity 62.7%; Pred. No. 2.26e-264;
Matches 222; Conservative 62; Mismatches 64; Indels 6; Gaps 6;
Db 1 MKRSISFTICLLITVTCGLQASPASAAAGTKTPAAKNQSLKIGTQLVNRDQKAVOLK 60
QY 1 MKKITTIFAV-LLMT-LALFSI-GNTTAADDYSV-VEEHGQLSISNGELVNERGEQVQLK 56
Db 61 GISSHGLQWYGFVYKESKWLKDDGKITVFRAMYTADGGYIDNPSVKNKVEAEAAK 120
QY 57 GMSHGLQWYGFVYKESKWLKDDGKITVFRAMYTSSGGYIDDPVSKVKETVEAAI 116
Db 121 ELGIYVIDWHILNDGNPNQKAKDFKEMSSLYGNTPNVIYEIANEPNG-DVNWKRD 179
QY 117 DLGIYVIDWHILNDGNPNYKKEAKDFDEMSELGYDNPVIYEIANEPNGSDVTWQNL 176
Db 180 IKPYAEVISVIRKNDPNIIIVGTGTSQVDAADQDLKANDVYALHYFAGTHGQSL 239
QY 177 IKPYAEVIVIRNDPNIIIVGTGTSQVDAADQDLKANDVYALHYFAGTHGQNL 236
Db 240 RDKANYALSKGAPFVTEWGTSDASNGGVFLDQSRWLNLYLDSKNISWYNWLSKQES 299
QY 237 RQVDYALDQGAIFVSEWGTSAATGGVFLDQSRWLNLYLDSKNISWYNWLSKQES 296
Db 300 SSALKPGASKTGWPLTDLTASGTFVRENIRGKDSKDPETPAQDNPTQKEG 353
QY 297 SAALMPGANPTGWTAEELSPSGTFVREKIRASIPSPDP-TPPSDPGEPDPG 349
RESULT 6
ENTRY #type complete
TITLE cellulase (EC 3.2.1.4) bg1C precursor - Bacillus subtilis
ALTERNATE_NAMES #formal_name Bacillus subtilis
ORGANISM #sequence_revision 05-Dec-1997 #text_change
DATE 05-Dec-1998
ACCESSIONS G69593; I40353; S24239; S49103; I39803
REFERENCE A69580
#authors
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Dueterhoeft, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galleron, N.; Gim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
```

V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tononi, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weltzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
#cross-references EMBL:G69593
#accession G69593
#molecule_type DNA
#status nucleic acid sequence not shown; translation not shown
#residues 1-508 #label KUN
#cross-references GB:299113; GB:AL009136; NID:g2634090; PID:e1183471; PID:g2634196
#experimental_source strain 168
REFERENCE A26114
#authors Mackay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moraneli, F.; Seligy, V.
#journal Nucleic Acids Res. (1986) 14:9159-9170
#title Structure of a *Bacillus subtilis* endo-beta-1,4-glucanase gene.
#cross-references EMBL:87066783
#accession A26114
#molecule_type DNA
#residues 10-508 #label MAC
#experimental_source strain PAP115
#note mature form, was confirmed by peptide sequencing

REFERENCE I40353
#authors Lindahl, V.; Aa, K.; Tronsmo, A.
#journal Antonie Van Leeuwenhoek (1994) 66:327-332
#title Nucleotide sequence of an endo-beta-1,4-glucanase gene from *Bacillus subtilis* CK-2.
#cross-references EMBL:95225656
#accession I40353
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 10-291, 'N', 293-508 #label LIN1
#cross-references EMBL:X67044; NID:g39776; PID:g39777
REFERENCE S24239
#authors Lindahl, V.; Aa, K.
#submission submitted to the EMBL Data Library, June 1992
#accession S24239
#molecule_type DNA
#residues 10-291, 'N', 293-508 #label LIN2
#cross-references EMBL:X67044; NID:g39776; PID:g39777
#experimental_source strain CK-2
REFERENCE S49103
#authors Wolf, M.; Geczi, A.; Bortiss, R.
#submission submitted to the EMBL Data Library, December 1993
#description Genes encoding beta glucan-hydrolyzing enzymes in *Bacillus subtilis*: construction of strains deficient in lichenase and cellulase activity.
#accession S49103
#molecule_type DNA
#residues 10-508 #label WOI
#cross-references EMBL:Z39076; NID:g509266; PID:g509267
REFERENCE I39803
#authors Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.
#journal Korean J. Microbiol. (1986) 24:236-242
#title Analysis on the nucleotide sequence of the signal region of *Bacillus subtilis* extracellular cellulase gene.
#accession I39803

##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'F', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PQ', 36-38, 'V', 40, 'E', 42-45, 'SI', 48-58, 'I', 60, 'Q', 62-80, 'DF', 83-86, 'T', 88-103, 'I', 105-157 #label SEO
##cross-references GB:M38634; NID:g142657; PID:g142658
##experimental_source strain ATCC 6633
COMMENT The low molecular weight of the mature protein suggests carboxyl-terminal processing as well as removal of the signal sequence.
GENETICS bglc
FUNCTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose and lichenin; can hydrolyze such linkages in beta-D-glucans that also contain 1,3-linkages
#pathway cellulose degradation
KEYWORDS extracellular protein; glycosidase; hydrolase; polysaccharide degradation
FEATURE 1-38
39-508
SUMMARY #domain signal sequence #status predicted #label SIG
#product cellulase #status predicted #label MAT
#length 508 #molecular-weight 56504 #checksum 4569
Query Match 48.0%; Score 1641; DB 2; Length 508;
Best Local Similarity 62.7%; Pred. No. 6.37e-263;
Matches 222; Conservative 60; Mismatches 66; Indels 6; Gaps 6;
Db 10 MKRSISFTICLLITLTMTGGMIA SPASAGTKTPVAKNGQLSKGTOLVNRDQKAVOLK 69
QY 1 MKKTIIFAV-LUMTL-ALFS-IGNTAADDSV-VEHQGLSISNGELVNERGEQVOLK 56
Db 70 GISSHGLQWYGEYVKNKSLKWLKRDGKITVFRAMTADGGYIDNPVSKNKEAVEAAK 129
QY 57 GMSHGLQWYGEYVKNKSLKWLKRDGKITVFRAMTADGGYIDNPVSKNKEAVEAAK 116
Db 130 ELGIYVVDHILNDGNPNONKEKKEFKEMSSLYGNTPNVIYEIANEPNG-DVNWKR 188
QY 117 DLGIYVVDHILNDGNPNONKEKKEFKEMSSLYGNTPNVIYEIANEPNG-DVNW 176
Db 189 IKPYAEVIVIRKNDPNDIIIVGTGTSQDVNDADQDLKANDVYALHFGYAGTHGQFL 248
QY 177 IKPYAEVIVIRKNDPNDIIIVGTGTSQDVNDADQDLKANDVYALHFGYAGTHGQ 236
Db 249 RKRYALSKGAPIFVTEWGTSDASNGGVFLDQSRWLKYLDSKTSYWNWNLSDQES 308
QY 237 RDQVDYALDQGAIFVSEWGTSAATGDDGVFLDEAQWIDFMDERNLISWANSLTHK 296
Db 309 SSALKPGASKTGWRSLDLSASGTFFVRENILGKSTKIDIPETPSKDKPTQENG 362
QY 297 SAALMPCANPTGCTEALSPSGTFVREKRESASIPPSDP-TPPSDPGPDGP 349

RESULT 7
ENTRY I40548 #type complete
TITLE bifunctional cellulase precursor - *Bacillus* sp.
ORGANISM #formal_name *Bacillus* sp.
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 17-Mar-1999
ACCESSIONS I40548
REFERENCE I40548
#authors Han, S.J.; Yoo, Y.J.; Kang, H.S.
#journal J. Biol. Chem. (1995) 270:26012-26019
#title Characterization of a bifunctional cellulase and its structural gene: the cel gene of *Bacillus* sp. D04 has exo- and endoglucanase activity.
#cross-references EMBL:96029707
#accession I40548
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-486 #label RES

Db 35 VETHGQLSIENGRVLVDEQKGKRVQLRGISSNGLQWVG DYVNKDSMKWL RDDWGINVFRVAM 94

מח

ELCTVTTDWHII NDCNPNONKFKAKFEFEKEMSSI VCNTPDNITVETANEBC - QUNWZDD 170

DB 121 ELCTVTTDRWHII NDCNBNONKEVAK EEEKFMSSY VCNTPDNVTVEYANEDNC - DVANRKPBD 170

Qy 3

RESULT 12

```

ENTRY          S29044      #type complete
TITLE          endoglucanase A precursor - Butyrivibrio fibrisolvens
ORGANISM       #formal_name Butyrivibrio fibrisolvens
DATE           13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
ACCESSIONS     S29044
REFERENCE      S29044
#authors       Hazlewood, G.P.; Davidson, K.; Laurie, J.I.; Romaniec,
                M.P.M.; Gilbert, H.J.
#journal       J. Gen. Microbiol. (1990) 136:2089-2097
#title         Cloning and sequencing of the cels gene encoding
                endoglucanase A of Butyrivibrio fibrisolvens strain A46.
#accession     S29044
#status        Preliminary
#molecule_type DNA
#residues      1-429 #label HAZ
#cross-references EMBL:M37031; NID:g144154; PID:g144155
SUMMARY        #length 429 #molecular-weight 48858 #checksum 682

Query Match    25.9%; Score 884; DB 2; Length 429;
Best Local Similarity 43.7%; Pred. No. 4.93e-127;
Matches 131; Conservative 65; Mismatches 97; Indels 11; Gaps 8;

Db 114 GKLAVERGLVDAGDGHVLLMGVSTGHINVPYASAEIKSLRDTWGINVIRLAMYTSQ 173
QY 36 GQUSISGELVNERGEQVQKGMSSHGLQVGFVNTESMKWLRDDGIVFRAAMTSS 95

Db 174 YNGCYVAGKEQEKLDIIDDVAEADNDMVIIDWHTLMDADPNYKADAIQFFGEMV 233
QY 96 -GGY-I--DDPS-VKEKVEAEADLGIYVIDWHILSDNDPNYKEAKOFFDEMS 149

Db 234 RYKNDENVYICNPNP--DTTWN--VRRYANEVPIVRNVA--IILVTPKWTDLQ 289
QY 150 ELXGDPNVIYEIANEPGSDVTDNQIKPYAEVPIVRNDPNNVIVGTGWSQDVH 209

Db 290 SVLDKPLDFNIMYTFYAGTHHKAERNALRDALDGLPFISEYGLVDADGDNLEK 349
QY 210 HAADNQLADPNVAFHYAGTHGQNRQDQVYALDQGAALFVSEWGTSAATGGGVFLD 269

Db 350 EADYWDIMIRKEYGVSSCMNLSNKGSAAMINADCKLSDFTEEDLSAMWLIDQIS 409
QY 270 EAQWIDFM-DEPNLSWANSLSLTHKDESSAALMGANPTGWTAEELSPSGTFVREKIRE 328

RESULT 13
ENTRY      JC5487      #type complete
TITLE      cellulase (EC 3.2.1.4) precursor - Clostridium cellulovorans
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM     #formal_name Clostridium cellulovorans
DATE         07-Jul-1997 #sequence_revision 12-Sep-1997 #text_change
ACCESSIONS   JC5487; PC4333
REFERENCE     JC5487
#authors     Shewetta, S.A.; Ichi-ishi, A.; Park, J.S.; Liu, C.; Malburg,
                L.M.; Doi, R.H.
#journal     Gene (1996) 182:163-167
#title       Characterization of enfF, a gene for a non-cellulosomal
                Clostridium cellulovorans endoglucanase.
#cross-references MUID:97136706
#accession   JC5487
#molecule_type DNA
#residues    1-557 #label SHE1
#cross-references GB:U37056; NID:g1778708; PID:g1778709
#experimental_source strain ATCC 35296
#accession   PC4333
#molecule_type protein
#residues    30-37 #label SHE2
COMMENT      This enzyme plays a secondary role in cellulose degradation.
GENETICS
#gene        enfF
KEYWORDS     glycosidase; hydrolase; polysaccharide degradation
FEATURES
1-29        #domain signal sequence #status predicted #label SIG

```

```

30-557
SUMMARY      #length 557 #molecular-weight 60131 #checksum 6863

Query Match    24.9%; Score 850; DB 2; Length 557;
Best Local Similarity 41.3%; Pred. No. 4.96e-121;
Matches 133; Conservative 76; Mismatches 91; Indels 22; Gaps 13;

Db 63 LCDKRGPIQLRGSTHGLWFPFVVNNNAALSDNSNVIRLAMYVAEGGYATNPVS 122
QY 45 LVNERGEQVQKGMSSHGLQVGFVNTESMKWLRDDGIVFRAAMTSSGIVDDPS 104

Db 123 KOTVINGNYALANDMYVIVDMMNPGDPNASVSG-AQSPFENDISTLYPNKNIYEL 181
QY 105 KEKVEIETEAALDLGIYVIDWHILSDNDPN--IKKEAKDFDEMSELYGDPNVIYEI 162

Db 182 CNEPENGSGVTNDATGWAQVSKYATPIVQLLRNKGNIILVGNPFSQRDPDLAADNPI 241
QY 163 ANEPNGSD--VTWDN---QIKPYAEVPIVRNDPNNVIVGTGWSQDVHHAADNQL 216

Db 242 NDSNTMYSVHFYSGTNPITSTVTDNRDNAMSNVRYALNHGAAVFATWGTSLATGTGPLY 301
QY 217 ADPNVYAFHYAGTH--GQ-NL-RDQ-V-D--YALDQGAALFVSEWGTSAATGGGVFL 268

Db 302 AKADAWLDFLNGNNTSWCFSTSNKDEKAAALNSITSLDPSDKLWADNELTSGOYVRA 361
QY 269 DEAQWIDFMDERNLSWANWSLTHKDESSAAL--MPGANPTGG--WTEAELSPSGTFVRE 324

Db 362 RIKGAYATPVDPVT--NOPTAP 382
QY 325 KIRASATPPSDPTPPDPGEP 346

RESULT 14
ENTRY      C42360      #type complete
TITLE      cellulase (EC 3.2.1.4) ES precursor - Thermomonospora fusca
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM     #formal_name Thermomonospora fusca
DATE         24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change
ACCESSIONS   C42360
REFERENCE     N42360
#authors     Lao, G.; Ghangas, G.S.; Jung, E.D.; Wilson, D.B.
#journal     J. Bacteriol. (1991) 173:3397-3407
#title       DNA sequences of three beta-1,4-endoglucanase genes from
                Thermomonospora fusca.
#cross-references MUID:91258320
#accession   C42360
#status      preliminary
#residues    1-466 #label LAO
#cross-references GB:L01577; NID:g154693; PID:g154694
FUNCTION      #description hydrolysis of 1,4-beta-D-glucosidic linkages in
                beta-D-glucans such as cellulose and lichenin; can
                hydrolyze such linkages in beta-D-glucans that also contain
                1,3-linkages
                cellulose degradation
#pathway      cellulose degradation
CLASSIFICATION #superfamily bacterial cellulose-binding domain homology
KEYWORDS       glycosidase; hydrolase; polysaccharide degradation
FEATURES
38-137        #domain bacterial cellulose-binding domain homology
                #label BCB
SUMMARY      #length 466 #molecular-weight 49760 #checksum 9976

Query Match    23.5%; Score 805; DB 2; Length 466;
Best Local Similarity 39.9%; Pred. No. 4.16e-113;
Matches 122; Conservative 81; Mismatches 88; Indels 15; Gaps 11;

Db 165 VERYKQVCGTQLCDEHGNPVQLRGMSTHGIQWFDHCLTDSLSALAYDWKADIRLSM 224
QY 32 VEEHQLSISNGELNERNERGEQVQKGMSSHGLQVGFVNTESMKWLRDDGIVFRAAM 91

Db 225 YIQEDGYETNPRGTFDRIDQLIDMATARGLYVIVDWHILTPGDPH-YNLDRAKTFPAETA 283

```


QY	92	YTSSGYYIDP-SVKEKVETVEAIDLGLYIWDHIIUSDNDPNLYK-EEARDFDEMS	149
Ddb	284	QRHASTKNVLYEIANEPNG--VSWAS-IKSYAEVIVPVRQRPDPSVITVGRSSSLGV	340
QY	150	ELYGDYPNVIYEIANEPNGSDVTWDMQIKPYAEVIVPVRDNDPNNVIVGTWSQ-DV	208
Ddb	341	SCSGSGPAETAANPVNASNTMYAFHFVAAASHRDNLNALREASEL-FPVVFEGFTWTG	399
QY	209	HHAAD-NOLA-DP---NVMYAFHFVAGTHGQNLRDQVDYALDQGNALFVSEWGSAA	262
Ddb	400	DGANDFQADRVIDLMAERKIGWTKWNYSDDFRSGAVFPGTCASGGPWSSSLKASQW	459
QY	263	DGVGFVDEAQVWIDFMDERNLSWANNLSLTHKDESSAALMPGANPTGG-WTEAELSPSGTF	321
Ddb	460	VRSKLQ	465
QY	322	VREKIR	327

```

RESULT      15
ENTRY
TITLE      S03767      #type complete
            cellulase (EC 3.2.1.4) z precursor - Erwinia chrysanthemi
            endo-1,4-beta-glucanase; extracellular endoglucanase z'
ORGANISM   #format_name Erwinia chrysanthemi
            strain 1937
DATE       21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change
            13-Sep-1998
ACCESSIONS S03767
REFERENCE   S03767
            #authors   Guiseppe, A.; Cami, B.; Aymeric, J.L.; Ball, G.; Creuzet, N.
            #journal    Mol. Microbiol. (1988) 2:159-164
            #title      Homology between endoglucanase z of Erwinia chrysanthemi and
                        endoglucanases of Bacillus subtilis and alkalophilic
                        Bacillus.
            #cross-references MUID:88216177
            #accession   S03767
                        ##molecule_type DNA
                        ##residues      1-428 ##label GUI
                        ##cross-references EMBL:Y00540; NID:g41091; PID:g41092
GENETICS
            #gene       celZ
FUNCTION
            #description hydrolysis of 1,4-beta-D-glucosidic linkages in
                        beta-D-glucans such as cellulose and lichenin; can
                        hydrolyze such linkages in beta-D-glucans that also contain
                        1,3-linkages
            #pathway      cellulose degradation
            #keywords      glycosidase; hydrolase; polysaccharide degradation
FEATURE
            1-43
            44-428
            ##domain signal sequence #status predicted #label SIG\
            #product cellulase z #status predicted #label MAT
            #length 428 #molecular-weight 46902 #checksum 683
SUMMARY

```

Query Match	23.2%	Score 793;	DB 2;	Length 428;
Best Local Similarity	40.1%;	Pred. No. 5,33e-111;		
Matches	125;	Conservative	69;	Mismatches 106; Indels 12; Gaps 10;
Ddb	76	WGGEFYTTADTVASLKDKWKSVIRAAAGVQSGGVLODPAGNKAKEVRVVDAATAANDWY	135	
Oy	65	WGQG-FVNYESKWLRRDDGIIIVFRAM-YTSSGGVIDPFSV-KEKVETVPAALDGLGY	121	
Ddb	136	AIGWHSHSAEN--RSEAIRFFOEMARKYGKNPNVIYEIYNEL--QVSWSNTIKPYA	190	
Oy	122	VIIDWHILSDNDPNHYKEEAKOFFDEMSELGDYPNVIEIANEPNGSDVTWDNQIKPYA	181	
Ddb	191	EAVISAIRAIDPDNLIIIVCTPSQSNVDSEASRDPLNAKNIATYTLFIYAGTHGESLURNKA	250	
Oy	182	EEVIPVRDNDRNNIVIGTGTSQDVHHAQNQLADPNVMTAFHFYAGTHGCQNLRQVD	241	
Ddb	251	QALNGIALFVTEGWIVNADGGNGVQNETDFAWTFMRDNNITQLTONMALNDKNEGASTY	310	
Oy	242	YALDGCAALFEVSEWTGTSATGDSGVFLFDQAQWIIDFMDBERNLSWA-NWSLTHKHDSSAAL	300	

```

Db      311 YPDSKNLRESKKVKSIQSWPYKAGSAASATWDPSTDTTMTPLNRRPOPTHQADCA 370
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      301 MPGA-NPT-GGWTEAELSPGCTFVREKIREASATPPSDPT-PPSDPCGPDGEPDTPPS 357
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      371 NANVYPNNVSKD 382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      358 DPGEYPAWDSNQ 369
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: Fri Jun 4 09:59:02 1999
Job time : 55 secs.

```

Search completed: Fri Jun 4 09:59:02 1999
Job time : 55 secs.

Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	2570	60.7	566	1	GUNB_PAELA	ENDOGLUCANASE B	PRECUR 0.00e+00
2	691	16.3	440	1	GUNB_CLOCL	ENDOGLUCANASE B	PRECUR 6.88e-11
3	673	15.9	515	1	GUND_CLOCL	ENDOGLUCANASE D	PRECUR 6.01e-11
4	660	15.6	814	1	GUNE_CLOMT	ENDOGLUCANASE E	PRECUR 4.17e-10
5	614	14.5	364	1	GUNA_RUMAL	ENDOGLUCANASE A (EC 3.	4.31e-10
6	603	14.2	409	1	GUNB_RUMAL	ENDOGLUCANASE B	PRECUR 1.05e-97
7	597	14.1	406	1	GUNL_RUMAL	ENDOGLUCANASE I	PRECUR 2.08e-96
8	525	12.4	473	1	GUNB_NEOPA	ENDOGLUCANASE B	PRECUR 6.43e-81
9	510	12.0	517	1	GUNA_CLOLO	ENDOGLUCANASE A	PRECUR 1.02e-77
10	489	11.5	547	1	GUNL_BUTFI	ENDOGLUCANASE 1 (EC 3.	2.96e-73
11	479	11.3	475	1	GUNA_CLOCE	ENDOGLUCANASE A	PRECUR 3.88e-71
12	422	10.0	584	1	GUND_CLOCE	ENDOGLUCANASE D	PRECUR 3.57e-59
13	412	9.7	900	1	GUNH_CLOCE	ENDOGLUCANASE H	PRECUR 4.28e-57
14	139	3.3	341	1	GUNL_CRYFL	ENDOGLUCANASE 1	PRECUR 1.76e-05
15	141	3.3	418	1	GUN2_TRELE	ENDOGLUCANASE EG-II	PR 8.72e-06
16	127	3.0	764	1	Y1ST_YEAST	HYPOTHETICAL 87.0 KD P	1.04e-03
17	118	2.8	336	1	GUNA_RUMFL	CELLODEXTRINASE A (EC	1.91e-02
18	111	2.8	1848	1	CBPA_CLOCL	CELLULOSE BINDING PROT	1.39e-02
19	112	2.6	264	1	PRIM_ENTFA	DNA PRIMASE (EC 2.7.7.	1.22e-01
20	104	2.5	181	1	YUAB_BAGSU	HYPOTHETICAL 19.3 KD P	1.29e+00
21	104	2.5	234	1	WAPA_BAGST	WALL-ASSOCIATED PROTEI	1.29e+00
22	105	2.5	304	1	MDH_HALMA	MALATE DEHYDROGENASE (9.69e-01
23	105	2.5	353	1	YBGO_ECOLI	HYPOTHETICAL 39.5 KD P	9.69e-01

Query Match 60.78; Score 2570; DB 1; Length 566;

```
-- - SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES). -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch)-----  
  
DB EML; M75706; G144792; --  
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.  
DR PROSITE; PS00448; CLOS_CELLULOSOME_RPT; 2.  
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.  
DR PFAM; PF00150; cellulase; 1.  
DR PFAM; PF00404; celcc; 2.  
DR HSP; P17901; IEDG.  
KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; REPEAT; SIGNAL.  
FT SIGNAL 1 33 POTENTIAL.  
FT CHAIN 34 440 ENDOGLUCANASE B.  
FT ACT_SITE 179 179 PROTON DONOR (BY SIMILARITY).  
FT ACT_SITE 305 305 NUCLEOPHILE (BY SIMILARITY).  
FT DOMAIN 387 440 2 X 24 AA APPROXIMATE REPEATS.  
FT REPEAT 387 410 1.  
FT REPEAT 419 440 2.  
SQ SEQUENCE 440 AA; 48637 MW; F1007BFC CRC32;  
  
Query Match 16.3%; Score 591; DB 1; Length 440;  
Best Local Similarity 34.0%; Pred. No. 6.88e-117;  
Matches 116; Conservative 98; Mismatches 103; Indels 24; Gaps 19;  
  
Db 13 LASVEVTTTFMGVVNVLASTAKTGIRDITSSQOVKEMKVGNLTGMTDATG-GETNWGP 71  
|||:::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
10 LAVLVVS-FVAFA-V--SSANDVKTLDIQSIVRDMPGVNLGTDFVGQDETAWGPN 65  
|||:::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
72 LTHAMDKYKAAGFNLRPLRTPDWDGHGAAPDAIDAATWNRVEEANTAFNNMVII 131  
I::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
66 RVTEELIERIADEGYKSIRIPVTWEINRIGGAPDPIDPQLNRVDDEVQALEDLYMI 125  
132 NLHHEDGWLPXY-YANE-AEVKAKITIKWTQIANRFKDYGDYLIIFETMNP RPVGAADEW 189  
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
126 NLHH-DSWLTIYMEHNNGVMARYSLRWELSQHFKDKPYTKLMFSVNPK-P--SQNV 181  
190 SSGSYENDDVNYRLNTAVNTIRATGGNNALRHIMVPTLAALAALSITMD-Y-IVPN-ND 246  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
182 GETREINHALLDDLDTVFIEVRISQGSDRIPLVLPTMETATSQPLLNNLYQTIDLDD 241  
247 SRVTSLSHYMSPFYSADLT SQMTATWGSDAADAASADFAYNKFKVNGRAVIGEM 306  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
242 PNLIATHVHYGFPPSFVNIA-G-YTR--FEESKR EIET-FDRVHTTVARGIPVVLGEF 297  
307 G-TINKNN-LD-S-RVKHAETYAKEATVRGITPIIWGDG 341  
I::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
298 GLLGFDKHKTGVIOOQEKLKFEEVLIHLNERDITHLWDNG 338  
  
RESULT 3  
ID Gund_Clocl STANDARD; PRt; 515 AA.  
AC P28623;  
DT 01-DEC-1992 (REL. 24, CREATED)  
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
DE 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE ENDOG_LUCANASE D PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE C)  
DN (CELLULOSE D).  
GN ENGd.  
OS CloSTRIDIUM CelluloVorans.  
OC BACTERIA; Firmicutes; Bacillus/CLOSTRIDIUM GROUP; CLoSTRIDIACEAE;  
OC CloSTRIDIUM.  
RN [1]  
RP Sequence from n.a., AND sequence of 32-44.  
PC STRAIN=atcc 35296;  
CY MEDIAN. 92167069
```

HAMAMOTO T., FOONG F., SHOSEYOV O., DOI R.H.;
 "Analysis of functional domains of endoglucanases from Clostridium
 cellulovorans by gene cloning, nucleotide sequencing and chimeric
 protein construction.";
 MOL. GEN. GENET. 231:472-479(1992).
 CC -!- FUNCTION: HAS ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYL-CELLULOSE
 CC (CMC), CELLOBIOSIDASE ACTIVITY ON P-NITROPHENYL-CELLOBIOSIDE
 CC (P-NPC), AND PARTIAL HYDROLYTIC ACTIVITY ON CRYSTALLINE CELLULOSE
 CC (AVICEL).
 CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 CC (CBD).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M37434; GI44796;
 CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC PFAM: PF00150; cellulase; 1.
 CC PFAM: PF00553; CBD.1; 1.
 CC HSP: P07986; IEXH.
 CC CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
 KW
 FT SIGNAL 1 31
 FT CHAIN 32 515 ENDOGLUCANASE D.
 FT DOMAIN 32 376 CATALYTIC (BY SIMILARITY).
 FT DOMAIN 377 407 PRO/THR-RICH (LINKER).
 FT DOMAIN 408 515 CELLULOSE-BINDING.
 FT ACT_SITE 180 180 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 303 303 NUCLEOPHILE (BY SIMILARITY).
 FT SEQUENCE 515 AA; 55976 MW; 268AAE53 CRC32;
 Query Match 15.9%; Score 673; DB 1; Length 515;
 Best Local Similarity 32.9%; Pred.No. 6.01e-113; Indels 11; Gaps 9;
 Matches 97; Conservative 90; Mismatches 97;
 Db 32 STAFTGVDRPAQOIVNEMKYGNLGMTDAIG-GETWGNPMTTHAMINKIKAGFNTL 90
 QY 24 SSANEDVKTLDIQSVVRDMPQGNLGNFTDVGQDETAWGNPRTRELIERIADGYSKI 83
 Db 91 RLPVTWGHGAPEYTDQTMKRVESIANAFDNDMVIINLHNEWLKPF-YANE- 148
 QY 84 RIPVTWENRIGGAPDYPIDPQFLNRVDEVQWALEEDLYVNLHDS-WLWYEMEHNY 142
 Db 149 AOYKAQLTKVQTQANFKKYGDLHIFETMNEPRPVGASLQWTGSGSYENRVNRYNLTA 208
 QY 143 NGVNAKTRSLVEQUNHFQDPTKLMPFESVNEPK-FSQN--WGEIRENHHALLDNLTVF 199
 Db 209 VNAIRATGGNNATRYIMVPTLAASAMSTINDL--VFN-NDKSVIVSLHMYSPYFFAMD 265
 QY 200 FEIVRQSGGQNDIRPLVPTMETATSPQLLNLTQTDKLDPNLIATVHYVGFSPFVYN 259
 Db 266 INGTSSGSDYDKSLSDDAVNFKNKGRVAVVIGEMSGINKNNTAARVTHAE 320
 QY 260 IAGYTFEEDSKREIET-FDRVHHTFVARGIPVVLGELGLGDKHTGVQOGE 313
 RESULT 4
 ID GUNE_CLOTH STANDARD; PRT; 814 AA.
 AC P10477;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
 DE 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE ENDOGLUCANASE E PRECURSOR (EC 3.2.1.4) (EGE) (ENDO-1,4-BETA-GLUCANASE)
 DE (CELLULOSE E).
 GN CELE.

OS CLOSTRIDIUM THERMOCCELLUM.
 CC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 CC CLOSTRIDIUM.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.
 RX MEDLINE: 89137992.
 RA HALL J., HAZLEWOOD G.P., BARKER P.J., GILBERT H.J.;
 "Conserved reiterated domains in Clostridium thermocellum
 RT endoglucanases are not essential for catalytic activity.";
 RL GENE 69:29-38(1988).
 CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
 CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
 CC GLUCANS.
 CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
 CC WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
 CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M22759; GI44770;
 CC PIR: JT0347; CZCLEM.
 DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PFAM: PF00150; cellulase; 1.
 DR PFAM: PF00404; celCC; 2.
 DR PFAM: PF00657; Lipase_GDSL; 1.
 DR HSP: P17901; IEDG.
 KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; REPEAT; SIGNAL.
 FT SIGNAL 1 34
 FT CHAIN 35 814 ENDOGLUCANASE E.
 FT ACT_SITE 193 193 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 316 316 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 415 474 2 X 24 AA APPROXIMATE REPEATS.
 FT REPEAT 415 438 1.
 FT REPEAT 451 474 2.
 FT SEQUENCE 814 AA; 90244 MW; AE61A167 CRC32;
 Query Match 15.6%; Score 660; DB 1; Length 814;
 Best Local Similarity 37.1%; Pred.No. 4.17e-110;
 Matches 104; Conservative 72; Mismatches 92; Indels 12; Gaps 10;
 Db 61 VKEIKIGNLGNLTDA--PTTAKGNPRTTKAMIEKVRMGFNNAVVPVTDTHIGPAD 118
 QY 39 VRDMQFNGNLGNTFDVAGQDETAWGNPRTRELIERIADGYSKIRIPVTWENRIGAPD 98
 Db 119 KYDEAWLNREVEVNYVLDCCGYALINLHDN-TWIIPTYANEQSKKLVKWEQIAT 177
 QY 99 YPIDQFLNRVDEVQWALEEDLYVNLHDSWLWYEMEHNYNGVMKYSRLWEQLSN 158
 Db 178 RPKDDHLLFFETMNEPRVSGSPMEWMTGTYENRD-VINRFLNAVVTIRASGNNDRKF 236
 QY 159 HPKDYPTKLMFESVNEPK-F-SQ-NW-GEIRENHHALLDNLTVFEEIVRQSGQNDIRP 214
 Db 237 ILVPTNAATGLDVALNDL--VIPN-NDKSVIVSIHAYSPIFFAMDVNGTSYMGSDYKAS 293
 QY 215 LVLPMTETATSPQLLNLTQTDKLDPNLIATVHYVGFSPVSNVAGYTFREED-SKRE 273
 Db 294 LSELDALYNRFVKNGRAVIGEGFTIDKNNLSSRVAHAE 333
 QY 274 IITFDRVHHTFVARGIPVVLGELGLGDKHTGVQOGE 313
 RESULT 5

ID	GU_NA_RUMAL	STANDARD;	PRT;	364 AA.
AC	P23660;			
DT	01-NOV-1991	(REL. 20, CREATED)		
DT	01-NOV-1991	(REL. 20, LAST SEQUENCE UPDATE)		
DT	01-JUN-1994	(REL. 29, LAST ANNOTATION UPDATE)		
DE	ENDOGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE)			
DE	(BGA).			
GN	CELA.			
OS	RUMINOCOCCUS ALBUS.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;			
OC	RUMINOCOCCUS.			
EN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.			
RC	STRAIN-SY3.			
RK	MEDLINE: 91066833.			
RA	POOLE D.M., HAZLEWOOD G.P., LAURIE J.I., BARKER P.J., GILBERT H.J.;			
RT	"Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase			
RT	genes celsa and celsb."			
RL	MOL. GEN. GENET. 223:217-223(1990).			
CC	- - FUNCTION: HYDROLYSES BOTH CARBOXYMETHYLCELLULOSE AND XYLAN.			
CC	PROBABLY HAS A ROLE IN HYDROLYZING OLIGOSACCHARIDES DERIVED			
CC	FROM CELLULOSE, WHICH ARE TRANSPORTED ACROSS THE CELL WALL.			
CC	- - CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC			
CC	LINKAGES IN CELLULOSE.			
CC	- - CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC			
CC	LINKAGES IN XYLANS.			
CC	- - SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).			
CC	- - SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X54931; G45964; "			
DR	PIR: S12017; S12017.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
DR	PFAM; PF00150; cellulase; 1.			
DR	HSPSP; P17901; LEDG.			
DR	CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; XYLAN DEGRADATION.			
FT	ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).			
FT	ACT_SITE 293 293 NUCLEOPHILE (BY SIMILARITY).			
SO	SEQUENCE 364 AA; 41218 MW; 680A4052 CRC32;			
	Query Match 14.5%; Score 614; DB 1; Length 364;			
	Best Local Similarity 35.8%; Pred. No. 4.31e-100;			
	Matches 101; Conservative 68; Mismatches 94; Indels 19; Gaps 16;			
Db	21 EVRDISAMELVGEMKGTGNLGNLSLDATGAPNASEVNGNPKTKTKEMIDAVYNGFDVR 80			
Qy	::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::			
Qy	29 DVKTLDIQSYVRDMPGPNLGNLTGDAVG-Q-D--ETAWGNPRVTRIELRIADEGYKSIR 84			
Db	81 IPVTWGGHVGDPADYKIDDEWIARVQEVVWYAYDDGAYVIINSHEEDW-RIPNEHI-D 138			
Qy	:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::			
Qy	85 IPTWENRIGAGPYPIDPQFLNRVDEVYQWALEEDYVMTNLHHD-SLWLIYEMEHYN 143			
Db	139 AVDEKTAIAIKQVAERFKDYGDHLIFEGNLEPRVYKSGPOEWNGCTEGRCVC-DRLNKTFF 197			
Qy	::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::			
Qy	144 GVMAYKYSRLWQLNSHFKDYPTKLMFSVNEPKF--S-QNW-GEIRENHALLDGLNTVF 199			
Db	198 LDTVRATGGNNKCR-LL--MTTYASSSM-SNVIKDTPAIPEDDDHGFSGSIHAYTPAFTYN 253			
Qy	::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::			
Qy	200 FEIVRQSGQNDIRPLVLPMTETATSOPLNLLNYQTIDKLDPNLIATVHYVGFWPFVS 259			
Db	254 ANADWELPHWDDSDHGGELVSLMTNLKENYLDKDPVITEYG 295			
Qy	:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::			
Qy	260 I-AGYTRFE-EDSKR-EIETTFDRVHHTFVARGIPVVLGEGF 298			

RESULT	6	STANDARD;	PRT;	409 AA.
ID	GUNB_RUMAL			
AC	P23661;			
DT	01-NOV-1991 (REL. 20, CREATED)			
DT	01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)			
DT	01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)			
DE	ENDOGLUCANASE B PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)			
DE	(CELLULOSE) (EGS).			
GN	CELB.			
OS	RUMINOCOCCUS ALBUS.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;			
OC	RUMINOCOCCUS.			
CC	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-SV3;			
RX	MEDLINE: 91066833.			
RA	POOLE D.M., HAZLEWOOD G.P., LAURIE J.I., BARKER P.J., GILBERT H.J.;			
RT	"Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase			
RT	genes celA and celB.;"			
RL	MOL. GEN. GENET. 223:217-223(1990).			
CC	-!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC			
CC	LINKAGES IN CELLULOSE.			
CC	-!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL:	X54932; G45966; -			
DR	PIR: S12018; S12018.			
DR	PROSITE; PS00013; PROKAR_LIPOPROTEIN; UNKNOWN_1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
DR	PFAM; PF00150; cellulase; 1.			
DR	HSP; P17901; LEDG.			
KW	CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.			
FT	SIGNAL			
FT	CHAIN			
FT	ACT_SITE 212 212 ENDOGLUCANASE B.			
FT	ACT_SITE 332 332 PROTON DONOR (BY SIMILARITY).			
FT	ACT_SITE 332 332 NUCLEOPHILE (BY SIMILARITY).			
SEQ	SEQUENCE 409 AA; 45523 MW; 777917FF CRC32;			
Query Match	14.2%; Score 603; DB 1; Length 409;			
Best Local Similarity	29.4%; Pred. No. 1.05e-97;			
Matches	104; Conservative 100; Mismatches 125; Indels 25; Gaps 21;			
Db	55 LPVSOTHTNDPMYTSAKDLVAKMSNGNGLNTMDATGEGLESEISMLPTKVYTNKFMID 114			
Qy	19 VAPAVSGANEDVKTLDTQSVYRDMQPGNGLNTFDVQGV--D-ETANGNPRV-TRE-LIE 73			
Db	115 MLPAGEFNVLRIPIYSWGNHLIDN-NYITDPAWMDRQVEIYNGIDDGMVILNTHHEW- 172			
Qy	74 RIADGYSIRIPVTWENRIGGADYPIDPQFLNRVDEVVQWALEEDLYVINLHDSWL 133			
Db	173 YM-KPPEKOGDIEELKAISQIADRPKGYDEHLIFPLNEPLRUGGEAWTGTSEARE- 230			
Qy	134 WIYEMHNINGVMARYKISLEQLSNHFHKDYPTKLMFESVNEPKF-SQ--NWGEIRENHA 190			
Db	231 LIINEVERAFVETVRASGNGNDCRLMI-T-GYAASSY--NNL-SATELPEDSKLLISVH 286			
Qy	191 LLDLNTVFFEIVRQSGQNDIRPLVLPTMETATSQPLNNLYQTDIKLDDPN-LIATVH 249			
Db	287 AYLPSFALDTKGYDKYDPEDT-AIPTLFFSLNELFTSRDIPVIVGFGSMKNDIDRV 345			
Qy	250 YIGFWPFSVNIAGVTRREEDSKRIELTFDRVHHTFVARGIPVVLGFGGLGDKHTGVI 309			
Db	346 KCLDDYLGNAKYDIPCVW-WDNVARI--GNGENGLLNROEYDWYFPLKMDVFK 397			
Qy	310 Q-QGEKLI-KFFEYLIIHLNPERDITLMLDNGOHF---NRHTYEWYDELFDMRLR 358			

```

RESULT 7
ID GUN1_RUMAL STANDARD; PRT; 406 AA.
AC F16216;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE I PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE) (EG-I).
GN EG I.
OS RUMINOCOCCUS ALBUS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC RUMINOCOCCUS.
RN [1]
SEQUENCE FROM N.A., AND SEQUENCE OF 44-58.
RC STRAIN=F-40;
RX MEDLINE; 90078126.
CC OHMIYA K., KAJINO T., KATO A., SHIMIZU S.;
CC "Structure of a Ruminococcus albus endo-1,4-beta-glucanase gene.";
CC J. BACTERIOL. 171:6771-6775(1989).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M30928; G152639; -.
DR DIR: A43722; A43722.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PFAM; PF00150; cellulase; 1.
DR HSP; P17901; 1EDG.
DR CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
FT SIGNAL 1 43
FT CHAIN 44 406 ENDOGLUCANASE I.
FT ACT_SITE 210 210 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 330 330 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 406 AA; 45390 MW; 5F3342B5 CRC32;

Query Match 14.1%; Score 597; DB 1: Length 406;
Best Local Similarity 31.6%; Pred. No. 2.08e-96;
Matches 92; Conservative 85; Mismatches 96; Indels 18; Gaps 14;

Db 53 VPVSTQHTNTMTVTSAKDLVAKTWGNLGNLTMDATAQGLGSVSVPLPKVTKNYKMD 112
: : : : : | | | | | | | | | | : : : : :
QY 19 VAPVSSANEDVKTLDTQSYVRDQPGNLTGNTDVGQD---ETANGNPRVTR--LIE 73
: : : : : | | | | | | | | | | : : : : :

Db 113 MLPAGFNVLRIPIVSGNHIIDQ-KYTSDPAMRDQVEIVNGIDNGLYILNTHHEW- 170
: : : : : | | | | | | | | | | : : : : :
QY 74 RIADGKYKSIRIPVTWENRIGGAPDYPIDQFLNKRVDVEVQWALEEDLYVMNLHDSWL 133
: : : : : | | | | | | | | | | : : : : :

Db 171 YM-PKPEKGDGDIIEIKAVWAQIADRPKGVDEHLIFGLNEPRLRGSAETGTSEARE- 228
: : : : : | | | | | | | | | | : : : : :
QY 134 WIYEMERNYNGVMKARYKLEQSLNHFKDYPKTKLMFVSNPKF-SQ--NWGEIRENHH 190
: : : : : | | | | | | | | | | : : : : :

Db 229 IINYEKAFVETVRASGNNGDRCLMT-T-GYAASSAY-NNL-SAIPEPDSDDLITSVH 284
: : : : : | | | | | | | | | | : : : : :
QY 191 LLDLNTVFEIVRQSGQNDIRPLVLPMTETATSQPLNNLYOTIDKLDDPN-LIATVH 249
: : : : : | | | | | | | | | | : : : : :

Db 285 AYLPSFALDTKGDYKDPEDT-AIPELFEHLNFIKGPVIVGFGTM 334
: : : : : | | | | | | | | | | : : : : :
QY 250 YYGFWPFSVNIAGYTRFEEDSKREIETFDVRVHHTFVARGIPVVLGEFGL 300
: : : : : | | | | | | | | | | : : : : :

RESULT 8
ID GUNB_NEOPA STANDARD; PRT; 473 AA.

```

```

AC Q12647;
AD
AE
AF
AG
AH
AI
AJ
AK
AL
AM
AN
AO
AP
AQ
AR
AS
AT
AU
AV
AW
AX
AY
AZ
BA
BB
BC
BD
BE
BF
BG
BH
BI
BJ
BK
BL
BM
BN
BO
BP
BQ
BR
BS
BT
BU
BV
BW
BX
BY
BZ
CA
CB
CC
CD
CE
CF
CG
CH
CI
CJ
CK
CL
CM
CN
CO
CP
CQ
CR
CS
CT
CU
CV
CW
CX
CY
CZ
DA
DB
DC
DD
DE
DF
DG
DH
DI
DJ
DK
DL
DM
DN
DO
DP
DQ
DR
DS
DT
DU
DV
DW
DX
DY
DZ
EA
EB
EC
ED
EE
EF
EG
EH
EI
EJ
EK
EL
EM
EN
EO
EP
EQ
ER
ES
ET
EU
EV
EW
EX
EY
EZ
FA
FB
FC
FD
FE
FF
FG
FH
FI
FJ
FK
FL
FM
FN
FO
FP
FQ
FR
FS
FT
FU
FV
FW
FX
FY
FZ
GA
GB
GC
GD
GE
GF
GG
GH
GI
GJ
GK
GL
GM
GN
GO
GP
GQ
GR
GS
GT
GU
GV
GW
GX
GY
GZ
HA
HB
HC
HD
HE
HF
HG
HH
HI
HJ
HK
HL
HM
HN
HO
HP
HQ
HR
HS
HT
HU
HV
HW
HX
HY
HZ
IA
IB
IC
ID
IE
IF
IG
IH
II
IJ
IK
IL
IM
IN
IO
IP
IQ
IR
IS
IT
IU
IV
IW
IX
IY
IZ
JA
JB
JC
JD
JE
JF
JG
JH
JI
JJ
JK
JL
JM
JN
JO
JP
JQ
JR
JS
JT
JU
JV
JW
JX
JY
JZ
KA
KB
KC
KD
KE
KF
KG
KH
KI
KJ
KK
KL
KM
KN
KO
KP
KQ
KR
KS
KT
KU
KV
KW
KX
KY
KZ
LA
LB
LC
LD
LE
LF
LG
LH
LI
LJ
LK
LL
LM
LN
LO
LP
LQ
LR
LS
LT
LU
LV
LW
LX
LY
LZ
MA
MB
MC
MD
ME
MF
MG
MH
MI
MJ
MK
ML
MM
MN
MO
MP
MQ
MR
MS
MT
MU
MV
MW
MX
MY
MZ
NA
NB
NC
ND
NE
NF
NG
NH
NI
NJ
NK
NL
NM
NO
NP
NQ
NR
NS
NT
NU
NV
NW
NX
NY
NZ
OA
OB
OC
OD
OE
OF
OG
OH
OI
OJ
OK
OL
OM
ON
OO
OP
OQ
OR
OS
OT
OU
OV
OW
OX
OY
OZ
PA
PB
PC
PD
PE
PF
PG
PH
PI
PJ
PK
PL
PM
PN
PO
PP
PQ
PR
PS
PT
PU
PV
PW
PX
PY
PZ
QA
QB
QC
QD
QE
QF
QG
QH
QI
QJ
QK
QL
QM
QN
QO
QP
QQ
QR
QS
QT
QU
QV
QW
QX
QY
QZ
RA
RB
RC
RD
RE
RF
RG
RH
RI
RJ
RK
RL
RM
RN
RO
RP
RQ
RR
RS
RT
RU
RV
RW
RX
RY
RZ
SA
SB
SC
SD
SE
SF
SG
SH
SI
SJ
SK
SL
SM
SN
SO
SP
SQ
SR
SS
ST
SU
SV
SW
SX
SY
SZ
TA
TB
TC
TD
TE
TF
TG
TH
TI
TJ
TK
TL
TM
TN
TO
TP
TQ
TR
TS
TT
TU
TV
TW
TX
TY
TZ
UA
UB
UC
UD
UE
UF
UG
UH
UI
UJ
UK
UL
UM
UN
UO
UP
UQ
UR
US
UT
UU
UV
UW
UX
UY
UZ
VA
VB
VC
VD
VE
VF
VG
VH
VI
VJ
VK
VL
VM
VN
VO
VP
VQ
VR
VS
VT
VU
VV
VV
VW
VX
VY
VZ
WA
WB
WC
WD
WE
WF
WG
WH
WI
WJ
WK
WL
WM
WN
WO
WP
WQ
WR
WS
WT
WU
WV
WW
WX
WY
WZ
XA
XB
XC
XD
XE
XF
XG
XH
XI
XJ
XK
XL
XM
XN
XO
XP
XQ
XR
XS
XT
XU
XV
XW
XX
XY
XZ
YA
YB
YC
YD
YE
YF
YG
YH
YI
YJ
YK
YL
YM
YN
YO
YP
YQ
YR
YS
YT
YU
YV
YW
YX
YY
YZ
ZA
ZB
ZC
ZD
ZE
ZF
ZG
ZH
ZI
ZJ
ZK
ZL
ZM
ZN
ZO
ZP
ZQ
ZR
ZS
ZT
ZU
ZV
ZW
ZX
ZY
ZZ

```

Db 291 VIIEGAM--NRNN--ED-DRERAEYIKKATSIGVPCVINDG-YFE 334
 QY 292 VVIEGFLGDKHTGVIQOGEKLFEXLIHHLNERDITHMLWDNQHFN 342

RESULT 9
 ID GUNA_CLOLO STANDARD; PRT; 517 AA.
 AC P54937;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE ENDOGLUCANASE A PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE A)
 DE (CELLULOSE A).
 GN CELA.

OS CLOSTRIDIUM LONGISPORUM.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 OC CLOSTRIDIUM.

RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN-ATCC 49440;

RX MEDLINE; 94172316.

RA MITTENDORF V., THOMSON J.A.;

RT "Cloning of an endo-(1->4)-beta-glucanase gene, celsa, from the rumen bacterium Clostridium sp. ('C. longisporum') and characterization of its product, Celsa, in Escherichia coli.";

RL J. GEN. MICROBIOL. 139:3233-3242(1993).

CC -!- FUNCTION: THE PH AND TEMPERATURE OPTIMA OF CELA ARE 4.8 AND 43 DEGREES CELSIUS, RESPECTIVELY. IT HYDROLYSES BARLEY BETA-GLUCAN, LICHENAN, CARBOXYMETHYLCELLULOSE AND XYLAN. IT SHOWS PREFERENTIAL ACTIVITY AGAINST THE LARGER CELLOOLIGOSACCHARIDES (CELLOHEXAOSE AND CELLOPENTAOSE); CELLOTetraose IS THE SMALLEST SUBSTRATE DEGRADED COMPLETELY.

CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; L02868; G144755; -

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

PFAM; PF00150; cellulase; 1.

PFAM; PF00553; CBD_1; 1.

DR HSSP; P17901; 1EDG.

KW CELLULOSE DEGRADATION: HYDROLASE; GLYCOSIDASE; SIGNAL.

FT SIGNAL 1 25

FT CHAIN 26 517

FT DOMAIN 26 ?

FT ACT_SITE 421 517

FT ACT_SITE 185 185

FT ACT_SITE 309 309

FT ACT_SITE 517 AA; 57660 MW; 8DEC293D CRC32;

Query Match 12.0%; Score 510; DB 1; Length 517;

Best Local Similarity 31.8%; Pred. No. 1.02e-77;

Matches 91; Conservative 85; Mismatches 89; Indels 21; Gaps 17;

Db 73 EFGWGNPVTKAMIDKIKNAGFKTIRIPITWGEHLDN-N-KLNEEWKRVKEVDYICIA 130

QY 59 ETAWGNPRVTRIELIERIADEGYKSIPIVTWENRIGGAPDYPIDPQFLNRVDEVQWALE 118

Db 131 DDLYVILNTHHG-NWVITYAKESVTPKLTWTQISEAFKDYDDHLIFELNPRLE 189

QY 119 EDLYVINLHDSWLWIYEMEHNYNGVMARYSLWQLSNHFKDPTKLMFESVNEPKF- 177

Db 190 GTPYEWTTGGTSESD-VVKNYNAALLESIRKTGGNNLSRAVMMPTYAASGSSITMND-FK 247
 QY 178 -SQ-NW-GEIRENHALLDLNTVFEIVRSGSQNDIRPLVLTMTATSQPLNLNYQ 234
 Db 248 -VP--DDKNVIASVHAYSPYFFAFMDTSSNSVNTWGSSYDYSLDVLDVLTNTFKSGVP 304
 QY 235 TIDKLDPNLIAVHYHGFVFPFVSVNIAGYT-R-FEESKRELIET-FDRVHTFVARGIP 291
 Db 305 VVIEGFG--SINKNNTS-SRAE-LA--EYVTAQKRGIPCVWMDN 344
 QY 292 VVIEGFLGDKHTGVIQOGEKLFEXLIHHLNERDITHMLWDN 337

RESULT 10

ID GUNL_BUTFI STANDARD; PRT; 547 AA.

AC P20847;

DT 01-FEB-1991 (REL. 17, CREATED)

DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)

DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)

DE ENDOGLUCANASE 1 (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE).

GN END1.

OS BUTYRIVIBRIO FIBRISOLVENS.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;

OC BUTYRIVIBRIO.

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-H17C;

RA MEDLINE; 90136507.

RT "Cloning and sequencing of an endoglucanase (endl) gene from Butyrivibrio fibrisolvens H17c.";

RL MOL. GEN. GENET. 219:193-198(1989).

CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X17538; G39473; -

DR PIR; JQ0356; JQ0356.

DR PROSITE; PS00561; CBD_BACTERIAL; 1.

DR PFAM; PF00659; GLYCOSYL_HYDROL_F5; 1.

DR PFAM; PF00150; cellulase; 1.

DR HSSP; P17901; 1EDG.

KW CELLULOSE DEGRADATION: HYDROLASE; GLYCOSIDASE.

FT ACT_SITE 189 189

FT ACT_SITE 321 321

FT DOMAIN 451 547

FT ACT_SITE 547 AA; 61078 MW; F318ABC3 CRC32;

Query Match 11.5%; Score 489; DB 1; Length 547;

Best Local Similarity 32.9%; Pred. No. 2.96e-73;

Matches 103; Conservative 84; Mismatches 94; Indels 32; Gaps 19;

Db 15 LALFVFMVMAIPATKVAAGGTDTRS-ATQV-VSDMRVGNWGNISLDSFGSQSYNFPYTSLE 72

QY 10 LAV-VLWSPVAPVSSANEDVKTLDIQSVVRDQPCWNLGNFTDVGQ-----DE 59

Db 73 TYGNCNPKATKALIDEVAKAGFNTRIPVSGQYTTGS-DYQI-PDFVMNRKVEVDYCV 130

QY 60 TAWGNPRVTRIELIERIADEGYKSIPIVTWENRIGGAPDYPIDPQF-LNRVDEVQWALE 118

```

CC CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC CC -!- THE C-TERMINUS (AA 411-475) MAY PLAY A ROLE IN ORGANIZING THE
CC CC CELLULOSE COMPLEX.
CC CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC CC HYDROLASES).
CC CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M93096; G144762; -
CC CC EMBL; M32362; G144759; -
CC CC PIR; JH0082; CZCLCA.
CC CC PDB; 1EDG; 17-AUG-96.
CC CC PROSITE; P800018; EF_HAND; UNKNOWN_1.
CC CC PROSITE; P800448; CLOS_CELLULOSE_RPT; 2.
CC CC PROSITE; P800659; GLYCOSYL_HYDROL_F5; 1.
CC CC PFAM; PF00150; cellulase; 1.
CC CC PFAM; PF00404; celCC; 2.
CC CC CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; REPEAT; SIGNAL;
CC CC 3D-STRUCTURE.
CC CC SIGNAL 1 26
CC CC CHAIN 27 475 ENDOGLUCANASE A.
CC CC ACT_SITE 147 147 PROBABLE.
CC CC ACT_SITE 195 195 PROTON DONOR.
CC CC ACT_SITE 332 332 NUCLEOPHILE.
CC CC DOMAIN 415 469 2 X 24 AA APPROXIMATE REPEATS.
CC CC REPEAT 415 438 1.
CC CC REPEAT 446 469 1.
CC CC MUTAGEN 104 104 R->K: SMALL LOSS OF ACTIVITY.
CC CC MUTAGEN 104 104 R->S,V: LARGE DECREASE OF ACTIVITY.
CC CC MUTAGEN 147 147 H->S,G,F: TOTAL LOSS OF ACTIVITY.
CC CC MUTAGEN 148 148 H->V: LARGE DECREASE OF ACTIVITY.
CC CC SEQUENCE 475 AA; 53625 MW; 3E9A8377 CRC32;
CC CC -----
Query Match 11.3%; Score 479; DB 1; Length 475;
Best Local Similarity 30.7%; Pred. No. 3.88e-71;
Matches 107; Conservative 96; Mismatches 110; Indels 35; Gaps 26;

Db 33 PNLIQIPQKNIPNDGNMFVKGILGNLGNLTFDAFGNTNITNEDLYETSGIKTTKQMI 92
| : : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 21 PAVSANEDYKTIIDIOSYVDMGPGWMLNLTGTFDA-----V-GQ-D-ETAWGNRVTRELI 72
| : : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 93 DALKQGFNTVRIPVSNHHPVSGS-DYKISDVMMNRVQEVVNYCIDNMKVILNTHHDVD 151
| : : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 73 ERTADGGYKIRIPVTWENRIGAPDIPDQPLNRVDEVVQWALEEDLVFMNLNHH-D-S 131

Db 152 KVKGYPPSSOYMASSKRYTITVWQAIAARFANYDEHLIFEGMNEPRLVGHANWPELTN 211
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 132 WLWIYEMEHNYNGVMARY-RSLWEOLSNHFKDYPYTKLMFESVNEPKF-SQ-N-WGEIRE 186

Db 212 SDVYDSINCINQLNQDFVNTVRATGKNASRYLMCPGY-VASPDGATNDYFRPNDSGN 270
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 187 NH--HAL--LDDLTNTVFEIVRSGSGGNDIRPLVPLTMTETATSQPLLNNLYQTIDKLD-D 241
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 271 NKKIIVSHAYCPWNFAGLAMADGGTNAWNINDSKQSEVTWMDNLYNKYTSRGIPVII 330
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 242 PN-LIATVHYGYFWPFS-VNIA-GYTR-FE-EDSKRIIET-F-DRVHHVTFVARGIPVL 294
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 331 GEGGAV--DKNN--LK-T-RVEYMSYVQAQAKARGILCILWDNN-NFS 371
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 295 GEGFLLGDKHTQVIQOGEKLFKEVYLHLHLNERDITHMLWDNGQHFN 342

RESULT 12
ID GUND_CLOCE STANDARD; PRT; 584 AA.
AC P25472;
DT 01-MAY-1992 (REL. 22. CREATED)

```


[illegible]

Query Match 3.3%; Score 139; DB 1; Length 341;
Best Local Similarity 20.6%; Pred. No. 1.76e-05;
Matches 22; Conservative 31; Mismatches 52; Indels 2; Gaps 2;

Db 62 EQVGHFTADGANLFRLPAGHQYLVGNNGQASTSLAPDFFAQYDALYQAVISKGAYAIIDVH 121
| : : : | : | : | : : : | : : : | : | : | : | : | : | : | : | : | : | :
QY 70 ELERIADEGYKIRIPVTWENRIGGA-PDPYPDPOFLNRVDVOWALEEDLYVMINLH 128
| : : : | : | : | : : : | : : : | : | : | : | : | : | : | : | : | : | :
Db 122 NYAR-WNGALIIGGGPSNQDFANLTLLATKVTSNDPNVFGLMNEP 167
| : : : | : | : | : : : | : : : | : | : | : | : | : | : | : | : | : | :
QY 129 HDSWLVIYEHHVNYGMKYRLWEQLSNHFKDYP TKLMFESVNEP 175
| : : : | : | : | : : : | : : : | : | : | : | : | : | : | : | : | : | :

RESULT 15

ID GUN2.TRIRE STANDARD; PRT; 418 AA.

AC P07982;

DT 01-AUG-1988 (REL. 08, CREATED)

DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE ENDOGLUCANASE EG-II PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
(CELLULASE).

DE EGL2 OR EGLII.

GN TRICHODERMA RESEI (HYPOCREA JECORINA).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;

QC HYPOCREALES; HYPOCREACEAE; HYPOCREA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VTT-D-80133;

RX MEDLINE; 8825850.

RA SALOHEIMO M., LEHTOVAARA P., PENTTILA M., TEERI T.T., STAHLBERG J., JOHANSSON G., PETTERSSON G., CLAYSSENS M., TOMME P., KNOWLES J.K.C.; "EgIII, a new endoglucanase from Trichoderma reesei: the characterization of both gene and enzyme."; GENE 63:11-21(1988). [2]

RN ACTIVE SITE GLQ-350.

RP MEDLINE; 93131031.

RX MACARON R., VAN BEEUMEN J., HENRISSAT B., DE LA MATA I., CLAESSENS M.;"Identification of an essential glutamate residue in the active site of endoglucanase III from Trichoderma reesei."; FEBS LETT 316:137-140(1993).

CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
(1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
(2) EXOCELLOBIOHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
(3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER SHORT CELLO-OLIGO-SACCHARIDES TO GLUCOSE.

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC -1- CAUTION: WAS ORIGINALLY CALLED ENDOGLUCANASE EG-III.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M19373; G170549; -.

DR PIR; S28372; S28372.

DR PROSITE; PS00562; CBD_FUNGAL; 1.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

DR PFAM; PF00150; cellulase; 1.

DR PFAM; PF00734; CBD_fungal; 1.

DR HSP; P00725; LAZH.

DW CELLULOSE DEGRADATION: HYDROLASE: GLYCOSIDASE: GLYCOPROTEIN: SIGNAL.

```

FT SIGNAL 1 21
FT CHAIN 22 418
FT DOMAIN 22 57
FT DOMAIN 58 91
FT DOMAIN 92 418
FT MOD_RES 22 22
FT CARBOHYD 124 124
FT DISULFID 29 46
FT DISULFID 40 56
FT ACT_SITE 239 239
FT ACT_SITE 350 350
SQ SEQUENCE 418 AA; 44227 MW; 7C18782A CRC32;

Query Match 3.3%; Score 141; DB 1; Length 418;
Best Local Similarity 22.8%; Pred. No. 8.72e-06;
Matches 23; Conservative 30; Mismatches 46; Indels 2; Gaps 2;

Db 142 VNEDGMTIFRLPVGMQVIVNNLGGNLDSTISIKYDQLVQGLSLGAYCIVDIHNYAR-W 200
QY 75 IADGYSIRIPVTWENRIGGAPDYPDPQFLNRVDEVVQWALEEDLYVMINLHDSWLW 134

Db 201 NGGIGGGPTNAQFTSLWSQLASKYAS-QSRVWFGIMNEP 240
QY 135 IYEMEHNYNGVMKYRSLWEQLSNHFKDYPYTKLMEESVNEP 175

```

Search completed: Fri Jun 4 10:01:32 1999
Job time : 31 secs.

WATERMAN

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPArch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 4 09:59:21 1999; MasPar time 18.98 Seconds
Tabular output not generated. 523.337 Million cell updates/sec

Title: >US-08-945-574-1
Description: (1-467) from US08945574.pep
Perfect Score: 3419
Sequence: 1 MKKITTIFAVLLMTLALFSI.....KWTQNEQPDGDPGWPEPLN 467
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part2 2:part3 3:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 35.687; Variance 195.452; scale 0.183

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	3419	100.0	467	19	W00382 Bacillus cellulase BC	1.17e-253
2	3419	100.0	467	20	W05731 Cellulase.	1.17e-253
3	3249	85.0	462	32	W57433 Closed alkaline endog	2.68e-240
4	2832	82.8	409	1	P81843 Sequence of alkaline	1.47e-207
5	2814	82.3	411	21	W12379 P300-CelB fusion cons	3.79e-206
6	2781	81.3	400	25	W23601 Bacillus agaradhaerens	1.47e-203
7	2781	81.3	400	32	W57431 Bacillus agaradhaerens	1.47e-203
8	2781	81.3	400	25	W25221 P300-CelB fusion cons	7.46e-203
9	2772	81.1	410	21	W12378 P300-CelB fusion cons	2.76e-197
10	2701	79.0	411	21	W12381 P300-CelB fusion cons	1.68e-196
11	2691	78.7	412	21	W12380 NK-1 cellulase.	4.80e-115
12	1649	48.2	499	8	R42122 Corrected Bacillus la	3.38e-108
13	1561	45.7	551	24	W18790 60 kD endoglucanase,	5.84e-91
14	1339	39.2	531	20	W01503 Endoglucanase encoded	3.48e-90
15	1329	38.9	532	3	R13229 Terebinibacter endogl	3.48e-54
16	862	25.2	1010	28	W34989	

17	759	22.2	956	36	W49874	Bankia gouldi glycosi	2.55e-46
18	739	21.6	822	5	R36021	Alkaline cellulase K-	8.48e-45
19	736	21.5	476	30	W37241	Heterodera glycines c	1.43e-44
20	733	21.4	800	2	P70420	Sequence encoded by c	2.42e-44
21	729	21.3	484	29	W43910	Heterodera glycines s	4.88e-44
22	696	20.4	472	30	W37243	Globodera rostochiens	1.57e-41
23	693	20.3	386	30	W37242	Globodera rostochiens	2.65e-41
24	690	20.2	357	15	R77394	Fragment of alkaline	4.47e-41
25	686	20.1	941	15	R77395	Full length Bacillus	4.47e-41
26	686	20.1	302	29	W43909	Heterodera glycines s	9.00e-41
27	670	19.6	941	2	R07478	Cellulase.	1.47e-39
28	626	18.3	319	30	W37240	Heterodera glycines c	3.12e-36
29	476	13.9	210	22	W17902	Cellulase from Bacill	5.49e-25
30	474	13.9	574	32	W57434	Termamyl-linker-CBD f	7.73e-25
31	400	11.7	130	30	W37245	Meloidogyne incognita	2.26e-19
32	398	11.6	130	30	W37246	Meloidogyne incognita	3.17e-19
33	396	11.6	130	30	W37244	Meloidogyne incognita	4.45e-19
34	382	11.2	560	25	W22522	Alpha-amylase-cellulo	4.71e-18
35	382	11.2	560	25	W23602	Alpha-amylase-cellulo	4.71e-18
36	215	6.3	106	29	W43911	Globodera rostochiens	3.69e-06
37	217	6.3	360	26	W43566	Thermotoga OC1/4V end	2.69e-06
38	217	6.3	360	36	W49870	Thermotoga OC1/4V end	2.69e-06
39	183	5.4	866	19	W02159	Soluble chitinase.	5.38e-04
40	171	5.0	517	9	R48671	Chitinase derivative	3.35e-03
41	139	4.1	826	5	R26042	P. yeellii SSP2 antige	3.86e-01
42	132	3.9	1046	19	W02156	Periplasmic chitoxide	1.06e+00
43	117	3.4	272	1	P93560	Plasmodium berghei ci	8.68e+00
44	116	3.4	449	23	W18210	Cellulomonas fimi end	9.97e+00
45	111	3.2	1911	18	R99534	Dermatomyositis speci	1.97e+01

ALIGNMENTS

RESULT 1
ID W00382 standard; Protein; 467 AA.

AC W00382;

DT 31-JAN-1997 (first entry)
DE Bacillus cellulase BCE 103.

KW Cellulase; BCE 103; detergent; surfactant; laundry;

KW tensile strength; antipilling

OS Bacillus sp. Strain CBS 670.93.

PN 31-Dec-1998.

PF 26-APR-1996; E01755

PR 28-APR-1995; EP-201115

PR 12-MAR-1998; EP-201115

PA (GENV) GENENCOR INT INC.

PI Kottwitz B, Lenting HBM,

PI Van Solingen P, Weiss A;

DR WPI: 96-457624/49.

DR N-PSDB; T41848.

PT Cellulase with low ratio of tensile strength loss to antipilling

PT properties - used in detergent composition which provides

PT anti-greying, softening, anti-wrinkling and colour protection to

PT fabrics

PS Claim 9; Fig 3; 33pp; English

CC A novel cellulase (W00382); designated BCE 103, of alkali-tolerant

CC Bacillus sp. Strain CBS 670.93 has a tensile strength loss to

CC antipilling ratio below 1. It can be isolated from CBS 670.93

CC fermentation broth or expressed at high levels in transformed host

CC cells utilising an isolated gene sequence (T41848). Cellulase BCE

CC 103, and similarly isolated cellulase BCE 113 (see also W00383),

CC show good activity at alkaline pH, and can be used in laundry

CC detergent compns. to provide anti-greying, softening, anti-

CC wrinkling and colour protection to fabrics.

SC Sequence 467 AA;

Query Match 100.0%; Score 3419; DB 19; Length 467;

Best Local Similarity 100.0%; Pred. No. 1.17e-253;

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mkkittifavllmtlalfsigntaadgysvveehgqlsngelvngegvqlkgmss 60

|||||

QY 1 MKKITTIFAVLLMTLALFSIGNTTAADDYSVVEEHGQLSISNGELVNERGEQVLKGMSS 60
 Db 61 hglwygqfnyesmkwlrddwgitvfraamytssggiddpsvkvetveaaidlgi 120
 QY 61 HGLQWYGQFVNYESMKWLRRDDWGITVFRAAMYTSSGGIDDPSVKVKETVEAAIDLGI 120
 Db 121 yviidwhilsdndpnnykeekadffdeneselygdyvpnyielanepngsdvtdnqikpy 180
 QY 121 YVIIDWHILSDNDPNNYKEEKADFFDENESELYGDPNPVYIELANEPNGSDVTWQNIKPY 180
 Db 181 aeevipvirdndpnnyivgtgtsqdvhaadnqladpnmvmafhyagthgqnlrdqv 240
 QY 181 AEEVIPVIRDNDPNNVIVGTGTSQDVVHAADNQLADPNVMYAFHYAGTHGQNLRDQV 240
 Db 241 dyaldqgaafvsewgttsaagdgvgvfldeaqvwidfmdernlswanwslthkdessaal 300
 QY 241 DYALDQGAAFVSEWGTSAAGDGGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
 Db 301 mpganptggwteaelspgstfvrekiresasippsdtpspdpgepdpptpsdpg 360
 QY 301 MPGANPTGGWTEAELSPGSTFVREKIREASIPPSDTPSPDPGEPDPTPPSPDG 360
 Db 361 eypawdsnqiytnelvyhngqlwqakwtqngqepdpgpweplksdpsdgsdpdtpspd 420
 QY 361 EYPAWDSNQIYTNELVYHNGQLWQAKWWTQNGEPDPGPWEPLKSDPSDGSDEPTPPSD 420
 421 pgeypawdsnqiytnelvyhngqlwqakwtqngqepdpgpwepln 467
 QY 421 PGEYPAWDSNQIYTNELVYHNGQLWQAKWWTQNGEPDPGPWEPLN 467

RESULT 2

ID W5731 standard; Protein; 467 AA.
 AC W5731;
 DT 29-MAR-1997 (first entry)
 DE Cellulase.
 KW Cellulase; Bacillus; alkalophilic bacterium; surfactant;
 KW Stonewashing; biopolishing; fabric softener; depilling.
 OS Bacillus sp. strain CBS 670.93.
 FH Key
 FT Key Location/Qualifiers
 FT Peptide 1..26
 FT /note= "Signal peptide"
 FT EP-739982-A1
 PD 30-OCT-1998
 PF 30-OCT-1998
 PR 30-APR-1995; EP-201115
 PA (GEMV) GENENCOR INC.
 PI Lenting HBM, Maurer KH, Van Beckhoven RFWC;
 PI Van Solingen P, Weiss A;
 DR WPI; 96-478746/48.
 DR N-PSDB; T40008.
 PT Cellulase isolated from novel Bacillus species - useful in detergent compsns, fabric softeners and de-pilling compsns, exhibits reduced adsorption of enzyme to fabric.
 CC Claim 6; Page 12-14; 17pp; English.
 CC The sequence represents a cellulase from alkalophilic Bacillus sp. CBS 670.93, and has been isolated in recombinant form by screening of Escherichia coli clones in plasmid pTZ18R. The enzyme shows a delta-REM of at least 4 units, preferably at least 5 units, in the Anti Redispersion Test, strong depilling activity, fibre damage of less than 0.05 mU in the Fibre Damage Test, and adsorption of less than 15% in the Adsorption Test. The cellulase may be used in surfactant compositions, and in stonewashing, biopolishing, fabric softener and depilling treatment compositions (claimed). The enzyme does not accumulate on the fabric after repeated laundry washing, and therefore has no effect on fabric tensile strength.
 SQ Sequence 467 AA;
 Query Match 100.0%; Score 3419; DB 20; Length 467;
 Best Local Similarity 100.0%; Pred. No. 1.17e-253;
 Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 mkkittifavllmtlalfsignttaaddysvveehgqlsngelvnrggeqvlkgmss 60

QY 1 MKKITTIFAVLLMTLALFSIGNTTAADDYSVVEEHGQLSISNGELVNERGEQVLKGMSS 60
 Db 61 hglwygqfnyesmkwlrddwgitvfraamytssggiddpsvkvetveaaidlgi 120
 QY 61 HGLQWYGQFVNYESMKWLRRDDWGITVFRAAMYTSSGGIDDPSVKVKETVEAAIDLGI 120
 Db 121 yviidwhilsdndpnnykeekadffdeneselygdyvpnyielanepngsdvtdnqikpy 180
 QY 121 YVIIDWHILSDNDPNNYKEEKADFFDENESELYGDPNPVYIELANEPNGSDVTWQNIKPY 180
 Db 181 aeevipvirdndpnnyivgtgtsqdvhaadnqladpnmvmafhyagthgqnlrdqv 240
 QY 181 AEEVIPVIRDNDPNNVIVGTGTSQDVVHAADNQLADPNVMYAFHYAGTHGQNLRDQV 240
 Db 241 dyaldqgaafvsewgttsaagdgvgvfldeaqvwidfmdernlswanwslthkdessaal 300
 QY 241 DYALDQGAAFVSEWGTSAAGDGGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
 Db 301 mpganptggwteaelspgstfvrekiresasippsdtpspdpgepdpptpsdpg 360
 QY 301 MPGANPTGGWTEAELSPGSTFVREKIREASIPPSDTPSPDPGEPDPTPPSPDG 360
 Db 361 eypawdsnqiytnelvyhngqlwqakwtqngqepdpgpweplksdpsdgsdpdtpspd 420
 QY 361 EYPAWDSNQIYTNELVYHNGQLWQAKWWTQNGEPDPGPWEPLKSDPSDGSDEPTPPSD 420
 421 pgeypawdsnqiytnelvyhngqlwqakwtqngqepdpgpwepln 467
 QY 421 PGEYPAWDSNQIYTNELVYHNGQLWQAKWWTQNGEPDPGPWEPLN 467

RESULT 3

ID W57433 standard; Protein; 462 AA.
 AC W57433;
 DT 01-SEP-1998 (first entry)
 DE Cloned alkaline endoglucanase protein sequence.
 KW Starch; liquefaction; sweetener; enzyme hybrid; endoglucanase; enzyme;
 KW Cellulose binding domain; CBD; starch processing; alpha-amylase;
 KW Saccharification.
 OS Bacillus agaradherens.
 FH Key Location/Qualifiers
 FT Peptide 1..26
 FT /note= "signal peptide"
 FT Protein 27..326
 FT /note= "endoglucanase enzyme"
 FT Peptide 327..354
 FT /note= "to be used as a linker sequence"
 FT Domain 355..400
 FT /note= "cellulose binding domain"
 FT Peptide 401..416
 FT /note= "to be used as a linker sequence"
 FT Domain 417..462
 FT /note= "cellulose binding domain"
 FN W57433-A1.
 PD 30-OCT-1998
 PF 30-OCT-1998
 PR 31-DEC-1996; DK-001130
 PA (NOVO) NOVO-NORDISK-AS.
 PI Bisgaardfrantzen H, Bjornvad M, Pedersen S, Schulein M;
 DR WPI; 98-251283/22.
 DR N-PSDB; V29663.
 PT Liquefaction of starch for, e.g. production of sweeteners - comprises use of enzyme hybrids including cellulose binding domain for starch
 PS Example 3; Pages 51-53; 83pp; English.
 CC This is the cloned Bacillus agaradherens endoglucanase sequence. This is used in the construction of enzyme hybrids for liquefaction of starch.
 CC The enzyme hybrids contain amino acid sequences of alpha-amylase linked to a cellulose binding domain (CBD). The starch is liquefied by treating in aqueous medium, with such an enzyme hybrid. A recombinant expression vector comprising a construct containing isolated DNA encoding enzyme hybrids with amylolytic activity, promoter and stop signals can be used

to transform host cells for the production of the recombinant enzyme hybrids. The enzyme hybrids are useful in industrial starch processing especially for the production of sweeteners. Hybrid enzymes have altered affinity for substrate and increased activity, resulting in at least 1 of reduced calcium ion dependence, reduced formation of Maillard reaction products and reduced effect of alpha-amylase on subsequent saccharification.

SQ Sequence 462 AA;

Query Match 95.0%; Score 3249; DB 32; Length 462;

Best Local Similarity 95.5%; Pred. No. 2,68e-240;

Matches 446; Conservative 9; Mismatches 7; Indels 5; Gaps 1;

```

Db 1 mkkittifvllmtalfisgnttaadsvveehqglisngelvngeqqlkgmss 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MKKITTIFAVLLMTALFISGNTTAADDSVVEEHGQLSISNGELVNERGEQQLKGMS 60
Db 61 hglqwygqfvnyesmkwlrddwgitvfraamytssgyiddpsvkekeveaaidldi 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 HGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGYIDDPSVKEKVEAEAAIDLGI 120
QY 61 HGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGYIDDPSVKEKVEAEAAIDLGI 120
Db 121 yviidwhilsdndpnlykeakdffdemselvgdypnviyeianepngsdvtwngikpy 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 YVIIDWHILSDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 180
QY 121 YVIIDWHILSDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 180
Db 181 aeepivlirndndpnlykeakdffdemselvgdypnviyeianepngsdvtwngikpy 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 AEVIPVIRNDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 240
QY 181 AEVIPVIRNDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 240
Db 241 dvaldgaalfvsewtsaatgvgvfldeaqvwidfmdernlswanwslthkdessaal 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 DYALDGAALFVSEWTSAAATGCGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
QY 241 DYALDGAALFVSEWTSAAATGCGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
Db 301 mpgantggvteaelspggtfvrekiresasippsdtpspdpgepdp 355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 MPGANTGGVTEAELSPSGTFFVREKIREASIPPSDTPSPDGPDPDPDPDPDPDP 355
QY 301 MPGANTGGVTEAELSPSGTFFVREKIREASIPPSDTPSPDGPDPDPDPDPDPDP 355
Db 356 kypawdpnglytneivhngqlwqakwtgnqepgdpypgwpelksdpsgepdp 415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
356 KYPADPNGLYTNELVHNGQLWQAKWTGNQEPGDPYPGPWPELKS DPSGE PDP 415
QY 356 KYPADPNGLYTNELVHNGQLWQAKWTGNQEPGDPYPGPWPELKS DPSGE PDP 415
Db 416 pgeypawdpnglytneivhngqlwqakwtgnqepgdpypgwpelksdpsgepdp 462
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
421 PGEYPADPNGLYTNELVHNGQLWQAKWTGNQEPGDPYPGPWPELKS DPSGE PDP 462

```

RESULT 4

ID P81843 standard; protein; 409 AA.

AC P81843;

DT 10-DEC-1990 (first entry)

DE Sequence of alkaline phosphatase encoded by Bacillus sp. No. N-4

DE N-4ACMCase gene

KW Enzyme.

OS Bacillus sp. No. N-4.

PN J6226392-2

PD 24-NOV-1987

PF 16-MAY-1986; 111928.

PA 16-MAY-1986; JP-111928.

PI (RIKA) Rikagaku Kenkyusho.

DR WPI: 88-004544/01.

DR N-PSDB: n82302.

PT Deoxyribonucleic acid sequence coding cellulase -

PT is obt'd. from naturally occurring matter, and hybridises with the

PT DNA of N-4ACMCase gene

PS Disclosure; Fig 3, Page 479; 12pp; Japanese.

CC The cellulase gene derived from Bacillus sp. No. N-4 is capable of

CC producing specific alkaline cellulase. The DNA can be obt'd. from natural

CC sources or by partial synthesis, and can hybridise with DNA of N-4ACMCase

CC gene.

SQ Sequence 409 AA;

Query Match

82.8%; Score 2832; DB 1; Length 409;

Best Local Similarity 94.9%; Pred. No. 1.47e-207;

Matches 388; Conservative 12; Mismatches 5; Indels 4; Gaps 2;

```

pb 1 mkkittifvllmtalfisgnttaadsvveehqglisngelvngeqqlkgmss 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MKKITTIFAVLLMTALFISGNTTAADDSVVEEHGQLSISNGELVNERGEQQLKGMS 60
Db 61 hglqwygqfvnyesmkwlrddwgitvfraamytssgyiddpsvkekeveaaidldi 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 HGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGYIDDPSVKEKVEAEAAIDLGI 120
QY 61 HGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGYIDDPSVKEKVEAEAAIDLGI 120
Db 121 yviidwhilsdndpnlykeakdffdemselvgdypnviyeianepngsdvtwngikpy 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 YVIIDWHILSDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 180
QY 121 YVIIDWHILSDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 180
Db 181 aeepivlirndndpnlykeakdffdemselvgdypnviyeianepngsdvtwngikpy 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 AEVIPVIRNDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 240
QY 181 AEVIPVIRNDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 240
Db 241 dvaldgaalfvsewtsaatgvgvfldeaqvwidfmdernlswanwslthkdessaal 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 DYALDGAALFVSEWTSAAATGCGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
QY 241 DYALDGAALFVSEWTSAAATGCGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
Db 301 mpgantggvteaelspggtfvrekiresasippsdtpspdpgepdp 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 MPGANTGGVTEAELSPSGTFFVREKIREASIPPSDTPSPDGPDPDPDPDPDPDP 360
QY 301 MPGANTGGVTEAELSPSGTFFVREKIREASIPPSDTPSPDGPDPDPDPDPDPDP 360
Db 361 sdpgdypawdpnglytneivhngqlwqakwtgnqepgdpypgwpelksdpsgepdp 409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
357 SDPGDYPADPNGLYTNELVHNGQLWQAKWTGNQEPGDPYPGPWPELKS DPSGE PDP 409

```

RESULT 5

ID W12379 standard; Protein; 411 AA.

AC W12379;

DT 17-JUN-1997 (first entry)

DE P300-CelB fusion construct 2 polypeptide product.

DE Cellulase; xylanase; alkaline protease; P300-CelB.

OS Chimeric-Bacillus-licheniformis-ATCC-53926;

OS Chimeric-Bacillus-sp. N4 (ATCC 21833).

FT Key

FT peptide

FT 1..26

FT /label= Sig-peptide

FT /note= "hybrid between P300 (aa1-5) and

FT CelB (aa8-26) signal peptides"

FT protein

FT 27..411

FT /label= Mat.protein

FT /note= "mature CelB cellulase"

PN W09706181-AL

PD 70-FEB-1997

PF 08-AUG-1996; U12545

PR 10-AUG-1995; US-002106.

PR 07-AUG-1996; US-694346.

PA (HENK) HENKEL CORP.

PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;

DR WPI: 97-154208/14.

DR N-PSDB: T63227.

PT System for increased expression of cellulase and xylanase in

PT Bacillus - contains gene under control of elements from B.

PT licheniformis-alkaline-protease-gene

PS Disclosure; Fig 8, 8A, 8B; 37pp; English.

CC The polypeptide product (W12379) of P300-CelB fusion construct 2

CC (T63227) comprises a hybrid signal peptide, formed between the

CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)

CC alkaline protease and Bacillus sp. N4 CelB cellulase, and the

CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.

CC host cells of P300-CelB fusion constructs 1-4 (see also W12378,

CC W12380-81) provides a 10-40 fold improvement of prodn. of the

CC alkalophilic cellulase in comparison to expression of the native

CC gene, and an earlier start to the prodn. of enzyme in the

CC fermentation process.

SQ Sequence 411 AA;

Query Match 82.3%; Score 2814; DB 21; Length 411;
 Best Local Similarity 94.4%; Pred. No. 3.79e-206; Mismatches 6; Indels 4; Gaps 2;
 Matches 385; Conservative 13;

Db 4 kktitfvllmtlalfignttaaddsvveehgqlsngelvndrgepvqlkgmssh 63
 QY 2 KKTITFAVLLMTLALFSGNTTAADDY SVVEEHGQLSNGELVNERGEQVQLKGMSH 61
 Db 64 glwqygfvnyesmkwlrddwqinvfraamytssggviedpsvkeveaaidlgiy 123
 QY 62 GLQWYQGFVNYESMKWLRDDWGITVFRAMYTSSGGVIDDP SVKEKVKETVEAAIDIGIY 121
 Db 124 viidwhilsndndniykeekakdfdemselgydypnviyeianepngsdvtdnqikpya 183
 QY 122 VIIDWHILSNDNDNIYKEEKAKDFDEMSELYGDPNVIYEIANEPNGSDVTWQNIKPIYA 181
 Db 184 eevipirndpniiivgtgtsqdvhaadnqltdpnvmyafhfagthgqnrdqvd 243
 QY 182 EEVIPVRNDPNIIIVGTGTSQDVVHAADNQLADPNVMAFHFYAGTHGQNLRDQVD 241
 Db 244 yaldqgaafvsewgtseatgdgvgvfldeaqvwidfmdernlswanwslthkdessaal 303
 QY 242 YALDQGAAFVSEWGTSAATGCGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAALM 301
 Db 304 pgsapgtggwteaelspggtfvrekiresattpsdptpsdpgepdpgepdpptps 363
 QY 302 PGANPTGGWTEAELSPGTFVREKIREASITPPSDPTPPSDP--GEP--DPGEPTTPPS 357
 Db 364 dpdgydpawdpntiaytdeivvnhngqlwqakwtnqepgdpypgwpwepin 411
 QY 358 DPGEYPAWDSNQIYTNIEIVHNGQLWQAKWWTQNEPFGDPYGPWEPLK 405

RESULT 6
 ID W23601 standard; Protein; 400 AA.
 AC W23601;
 DT 08-FEB-1998 (first entry)
 DE Bacillus agaradherens alkaline cellulase Cel5A;
 KW Desizing; cellulose; fabric; enzyme hybrid; alkaline cellulase;
 OS endoglucanase; Bacillus agaradherens; cellulose binding domain;
 PN Cel5A.
 OS Bacillus agaradherens strain NCIMB 40482.
 PN W09728243-AL.
 PD 07-AUG-1997.
 PF 29-JAN-1997; DK0042.
 PR 29-JAN-1996; DK-000094.
 PA (NOVO) NOVO-NORDISK AS.
 PI Bjornvad ME, Cherry JR, Rasmussen MD, Vind J, Von Der Osten C;
 DR WPI: 97-402598/37.
 DR N-PSDB; T74270.
 PT Cleaning of cellulosic fabrics - using an enzyme hybrid comprising a
 sequence of a non-cellulolytic enzyme linked to a cellulose-binding
 domain sequence
 Claim 17; Page 95-96; 124pp; English.
 CC This protein comprises the alkaline cellulase (endoglucanase) of
 Bacillus agaradherens NCIMB 40482. It was expressed in Bacillus
 subtilis P2306 transformants following PCR amplification (see
 CC T74288-89) of B. agaradherens genomic DNA and ligation of the PCR
 CC product into vector pDN1981. DNA encoding the cellulose binding
 CC domain (CBD) of the alkaline cellulase was subsequently amplified
 CC (see T74290-91) for use in the construction of a novel alpha-
 CC amylase-CBD hybrid enzyme (see T74271). A claimed process for
 CC removal or bleaching of soiling or stains on a cellulosic fabric
 CC comprises contacting the fabric with a modified enzyme (enzyme
 CC hybrid) comprising a catalytically active portion of a non-
 CC cellulolytic enzyme linked to a CBD. The hybrid enzyme gives
 CC improved enzyme performance by increasing the affinity of the
 CC enzyme for the fabric.
 SQ Sequence 400 AA;

Query Match 81.3%; Score 2781; DB 25; Length 400;
 Best Local Similarity 95.6%; Pred. No. 1.47e-203;
 Matches 387; Conservative 8; Mismatches 5; Indels 5; Gaps 1;

Db 1 mkkittifvllmtlalfignttaaddsvveehgqlsngelvnergeqvqlkgmss 60
 QY 1 MKKITTIFAVLLMTLALFSGNTTAADDY SVVEEHGQLSNGELVNERGEQVQLKGMS 60
 Db 61 hglwqygfvnyesmkwlrddwqinvfraamytssggviedpsvkeveaaidlgi 120
 QY 61 HGLWYQGFVNYESMKWLRDDWGITVFRAMYTSSGGVIDDP SVKEKVKETVEAAIDIGI 120
 Db 121 viidwhilsndndniykeekakdfdemselgydypnviyeianepngsdvtdnqikpy 180
 QY 121 VIIDWHILSNDNDNIYKEEKAKDFDEMSELYGDPNVIYEIANEPNGSDVTWQNIKPY 180
 Db 181 aeevipirndpniiivgtgtsqdvhaadnqltdpnvmyafhfagthgqnrdqvd 240
 QY 181 AEEVIPVRNDPNIIIVGTGTSQDVVHAADNQLADPNVMAFHFYAGTHGQNLRDQV 240
 Db 241 yaldqgaafvsewgtseatgdgvgvfldeaqvwidfmdernlswanwslthkdessaal 300
 QY 241 YALDQGAAFVSEWGTSAATGCGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
 Db 301 mgsanptggwteaelspggtfvrekiresattpsdptpsdpgepdpdpdpdpdp 355
 QY 301 MGSANPTGGWTEAELSPGTFVREKIREASITPPSDPTPPSDPGEPTPPSDPG 360
 Db 356 eypawdpnqiytnieivvnhngqlwqakwtnqepgdpypgwpwepin 400
 QY 361 EYPAWSNQIYTNIEIVHNGQLWQAKWWTQNEPFGDPYGPWEPLK 405

RESULT 7
 ID W57431 standard; Protein; 400 AA.
 AC W57431;
 DT 01-SEP-1998 (first entry)
 DE Bacillus agaradherens endoglucanase enzyme.
 KW Starch; liquefaction; sweetener; enzyme hybrid; endoglucanase; enzyme;
 OS cellulose binding domain; CBD; starch processing; alpha-amylase;
 OS saccharification.
 OS Bacillus agaradherens.
 PN W09816633-AL.
 PD 23-APR-1998.
 PF 13-OCT-1997; DK0448.
 PR 11-OCT-1996; DK-001130.
 PA (NOVO) NOVO-NORDISK AS.
 PI Bisgardfrantzen H, Bjornvad M, Pedersen S, Schulein M;
 DR WPI: 98-251283/22.
 DR N-PSDB; V29654.
 PT Liquefaction of starch for, e.g. production of sweeteners -
 PT comprises use of enzyme hybrids including cellulose binding domain
 PT for starch
 PS Example 1; Pages 46-47; 83pp; English.
 CC This represents a Bacillus agaradherens endoglucanase enzyme. This is
 CC used in the construction of enzyme hybrids for liquefaction of starch.
 CC The enzyme hybrids contain amino acid sequences of alpha-amylase linked
 CC to a cellulose binding domain (CBD). The starch is liquefied by treating,
 CC in aqueous medium, with such an enzyme hybrid. A recombinant expression
 CC vector comprising a construct containing isolated DNA encoding enzyme
 CC hybrids with amylolytic activity, promoter and stop signals can be used
 CC to transform host cells for the production of the recombinant enzyme
 CC hybrids. The enzyme hybrids are useful in industrial starch processing
 CC especially for the production of sweeteners. Hybrid enzymes have altered
 CC affinity for substrate and increased activity, resulting in at least 1 of
 CC reduced calcium ion dependence, reduced formation of Maillard reaction
 CC products and reduced effect of alpha-amylase on subsequent
 CC saccharification.
 SQ Sequence 400 AA;

Query Match 81.3%; Score 2781; DB 32; Length 400;
 Best Local Similarity 95.6%; Pred. No. 1.47e-203;
 Matches 387; Conservative 8; Mismatches 5; Indels 5; Gaps 1;

Db 1 mkkittifvllmtlalfignttaaddsvveehgqlsngelvnergeqvqlkgmss 60
 QY 1 MKKITTIFAVLLMTLALFSGNTTAADDY SVVEEHGQLSNGELVNERGEQVQLKGMS 60

QY 1 MKKITTIFAVLLMTLALFSIGNTTAADDYSVVEEHGOLSISNGELVNERGEQVQLKGMSS 60
 Db 61 hglwqygqfnyesmkwlrddwginvfraamytssggylidpsvkekveaveaaidldi 120
 QY 61 HGLWYQGFVNYESMKWLRDDWGITVFRAMYTSSGGYIDDPVSKVKETVEAAIDLGI 120
 Db 121 yviidwhilsdndpnlykeekadffdemseelydypnvlyeianepngsdvtwgnqikpy 180
 QY 121 YVIIDWHILSDNDPNLYKEEAKDFFDEMSELYGDYPNVLYEIANEPNGSDVTWGNQIKPY 180
 Db 181 aeepvpirndpnllivgtgtsqdvvhhaadnqladpnmvafhyfagthgqnldqv 240
 QY 181 AEVIPVIRNDPNLLIVGTGTSQDVVHHAADNQLADPNMVAFHYFAGTHGQNLRDQV 240
 Db 241 dvaldgaalfvsewtsaatgvgvfldeaqvwidfmdernlswanwslthkdessaal 300
 QY 241 DYALDQGAALFVSEWTSAAATGGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
 Db 301 mpgantggwteaelspsgtfvrekiresasippsdtpspdpgepdp-----tppsdpq 355
 QY 301 MPGANTGGWTEAELSPSGTFVREKIREASIPSPDTPSPDPGEPDPTPPSPDPG 360
 Db 356 eypawdpnclytneivyhngqlwqakwttqngqgdpygpwepln 400
 QY 361 EYPAWSNQIYTNIEIVYHNGQLWQAKWTTQNGQGDYPGPWEPLK 405

RESULT

ID W22521 standard; Protein; 400 AA.
 AC W22521;
 DT 08-FEB-1998 (first entry)
 DE Bacillus agaradherens alkaline cellulase Cel5A;
 KW Desizing; Cellulose; fabric; enzyme hybrid; alkaline cellulase;
 KW endoglucanase; Bacillus agaradherens; cellulose binding domain;
 KW Cel5A.
 OS Bacillus agaradherens strain NCIMB 40482.
 PN W09728256-A1.
 PD 07-AUG-1997.
 PF 29-JAN-1997; DK0041.
 PR 29-JAN-1996; DK-000093.
 PA (NOVO) NOVO-NORDISK AS.
 PI Bjornvad ME, Rasmussen MD, Vind J, Von Der Osten C;
 DR WPI; 97-403610/37.
 PS N-PSDB; T77055.
 PT Desizing cellulose-containing fabric or textile using an enzyme
 hybrid - which comprises a catalytically active amino acid sequence
 of a non-cellulolytic enzyme linked to an amino acid sequence
 comprising a cellulose binding domain.
 PS Example 2; Page 52-54; 72pp; English.
 CC This protein comprises the alkaline cellulase (endoglucanase) of
 Bacillus agaradherens NCIMB 40482. It was expressed in Bacillus
 subtilis P22306 transformants following PCR amplification (see
 T77053-54) of B. agaradherens genomic DNA and ligation of the PCR
 product into vector pDN1981. DNA encoding the cellulose binding
 domain (CBD) of the alkaline cellulase was subsequently amplified
 (see T77056-57) for use in the construction of a novel alpha-
 amylase-CBD hybrid enzyme (see T77058). A claimed process for
 desizing cellulose-containing fabric or textile comprises treating
 the fabric or textile with a modified enzyme (enzyme hybrid)
 comprising a non-cellulolytic enzyme linked to a CBD. The process
 gives improved enzyme performance by modifying the enzyme so as to
 increase its affinity for cellulosic fabric.
 CC Sequence 400 AA;

Query Match 81.3%; Score 2781; DB 25; Length 400;
 Best Local Similarity 95.6%; Pred. No. 1.47e-203;
 Matches 387; Conservative 8; Mismatches 5; Indels 5; Gaps 1;

Db 1 mkkittifavllmtalvlsignttaadndsvveehgolsisngelvnrgsqvqlkqms 60
 QY 1 MKKITTIFAVLLMTALVLSIGNTTAADDYSVVEEHGOLSISNGELVNERGEQVQLKGMSS 60
 Db 61 hglwqygqfnyesmkwlrddwginvfraamytssggylidpsvkekveaveaaidldi 120

QY 61 HGLWYQGFVNYESMKWLRDDWGITVFRAMYTSSGGYIDDPVSKVKETVEAAIDLGI 120
 Db 121 yviidwhilsdndpnlykeekadffdemseelydypnvlyeianepngsdvtwgnqikpy 180
 QY 121 YVIIDWHILSDNDPNLYKEEAKDFFDEMSELYGDYPNVLYEIANEPNGSDVTWGNQIKPY 180
 Db 181 aeepvpirndpnllivgtgtsqdvvhhaadnqladpnmvafhyfagthgqnldqv 240
 QY 181 AEVIPVIRNDPNLLIVGTGTSQDVVHHAADNQLADPNMVAFHYFAGTHGQNLRDQV 240
 Db 241 dvaldgaalfvsewtsaatgvgvfldeaqvwidfmdernlswanwslthkdessaal 300
 QY 241 DYALDQGAALFVSEWTSAAATGGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
 Db 301 mpgantggwteaelspsgtfvrekiresasippsdtpspdpgepdp-----tppsdpq 355
 QY 301 MPGANTGGWTEAELSPSGTFVREKIREASIPSPDTPSPDPGEPDPTPPSPDPG 360
 Db 356 eypawdpnclytneivyhngqlwqakwttqngqgdpygpwepln 400
 QY 361 EYPAWSNQIYTNIEIVYHNGQLWQAKWTTQNGQGDYPGPWEPLK 405

RESULT

ID W12378 standard; Protein; 410 AA.
 AC W12378;
 DT 17-JUN-1997 (first entry)
 DE P300-CelB fusion construct 1 polypeptide product.
 KW Cellulase; xylanase; alkaline protease; P300; CelB.
 OS Chimeric Bacillus licheniformis ATCC 53926;
 OS Chimeric Bacillus sp. N4 (ATCC 21833).
 FH Key Location/Qualifiers
 FT peptide 1..26
 FT /label= Sig_peptide
 FT /note= "hybrid between P300 (aal-12) and
 CelB (aal3-26) signal peptides"
 FT protein 27..410
 FT /label= Mat_protein
 FT /note= "mature CelB cellulase"
 PN W03706181-A1.
 PD 20-FEB-1997.
 PF 08-AUG-1996; UI2545.
 PR 10-AUG-1995; US-002106.
 PR 07-AUG-1996; US-694346.
 PA (HENK) HENKEL CORP.
 PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
 DR WPI; 97-154208/14.
 DR N-PSDB; T63226.
 PT System for increased expression of cellulase and xylanase in
 Bacillus - contains gene under control of elements from B.
 PT licheniformis alkaline protease gene
 PS Disclosure; Fig 7,7A,7B; 37pp; English.
 CC The polypeptide product (W12378) of P300-CelB fusion construct 1
 (T63226) comprises a hybrid signal peptide, formed between the
 CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
 CC alkaline protease and Bacillus sp. N4 CelB cellulase, and the
 CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
 CC host cells of P300-CelB fusion constructs 1-4 (see also
 CC W12379-81) provides a 10-40 fold improvement of prodn. of the
 CC alkalophilic cellulase in comparison to expression of the native
 CC gene, and an earlier start to the prodn. of enzyme in the
 CC fermentation process.
 CC Sequence 410 AA;

Query Match 81.1%; Score 2772; DB 21; Length 410;
 Best Local Similarity 92.9%; Pred. No. 7.46e-203;
 Matches 380; Conservative 17; Mismatches 8; Indels 4; Gaps 2;

Db 2 mrksfslgmllmtalvlsignttaadndsvveehgolsisngelvnrgsqvqlkqms 61
 QY 1 MKKITTIFAVLLMTALVLSIGNTTAADDYSVVEEHGOLSISNGELVNERGEQVQLKGMSS 60

QY	1	MKKTTTFAVLLMTLAL-FSIGN-TTAAADYSVVEEHGOLSIISNGELVNERGEQVOLKGM	58
Db	61	sshgqlwygqfvnyeskmkrlrdwgtvfraamytsggyiedpsvkexkveaveaaigl	120
QY	59	SSHGLWYGQFVNYESKMWLRRDGGITVFRAAMYTSSGGYIDDPSPVKETVEAIDL	118
Db	121	giyvildwhllsdndpnlykeeakdfdemselfygdypnviyeianepngsdvtwdnqli	180
QY	119	GIVYIDWHILSDNDPNLYKEEAKDFDEMSELYGYPNVIYEIANEPNGSDVTWDNQIK	178
Db	181	pyaeeripvirndpnlliivtqtswsqvhhaadnqlldpnvmvafhyagthgnlrd	240
QY	179	PYAEEVIPVRINDPNNIVIGTGTWSQDVHHAADNQLDNVMYAFHFYAGTHGNLRD	238
Db	241	qvdyalddgaifsvewgtseatdgqgvfldeaqvwldfmdernlswnawstlhkdessa	300
QY	239	QVDYALDDGAARIFVSEWGTSATDCGGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSA	298
Db	301	almppgasptgwteaelspgstfvrekiresattppsdpdtpgsdpdpgepdpgedpdt	360
QY	299	ALMPGANPTGWTEAEELSPGSTFVREKIREASIPPSDPTPPSDP--GEP--DPGEPDPT	354
Db	361	ppsdgdyapawdntlytdeiynhngqlwgakwtcnqeqggdygpwpeln	411
QY	355	PPSDGGEYPAMDNSQIYTNIEIVYHNGLWQAQKWWTQNQPGDPGPWPPLK	405
RESULT 11			
ID	W12380	standard; Protein; 412 AA.	
AC	W12380;		
DT	17-JUN-1997	(first entry)	
DE	P300-CelB fusion construct 3 polypeptide product.		
KW	Cellulase; xylanase; alkaline protease; p300; CelB.		
OS	Chimeric Bacillus licheniformis ATCC 53926;		
OS	Chimeric Bacillus sp. N4 (ATCC 21833).		
FH	Key	Location/Qualifiers	
FT	peptide	1..29	
FT	/label= Sig_peptide	/label= "mature CelB cellulase"	
FT	/note= P300 signal peptide"		
FT	protein	30..412	
FT	/label= Mat_protein		
FT	/note= "mature CelB cellulase"		
FN	WO9706181-A1.		
PN	20-FEB-1997.		
PF	08-AUG-1996; U12545.		
PR	10-AUG-1995; US-002106.		
PR	07-AUG-1996; US-694346.		
PI	(HENK) HENKEL CORP.		
PA	Christianston T, Maurer K, Tang MR, Weiss A, Wilson CR;		
DR	WPI; 97-154208/14.		
DR	N-PSDB; T63228.		
PT	System for increased expression of cellulase and xylanase in		
PT	Bacillus - contains gene under control of elements from B.		
PT	Licheniformis alkaline protease gene		
PS	Disclosure; Fig 9,9A,9B; 37pp; English.		
CC	The polypeptide product (W12380) of p300-CelB fusion construct 3		
CC	(T63228) comprises the signal peptide of Bacillus licheniformis		
CC	ATCC 53926 (P300) alkaline protease and the mature CelB		
CC	alkalophilic cellulase of Bacillus sp. N4. Expression in Bacillus		
CC	sp. host cells of P300-CelB fusion constructs 1-4 (see also W12378-		
CC	79, W12381) provides a 10-40 fold improvement of prodn. of the		
CC	alkalophilic cellulase in comparison to expression of the native		
CC	gene, and an earlier start to the prodn. of enzyme in the		
CC	fermentation process.		
SQ	Sequence 412 AA;		
Query Match 78.7%; Score 2691; DB 21; Length 412;			
Best Local Similarity 90.3%; Pred. No. 1.68e-196;			
Matches 364; Conservative 28; Mismatches 6; Indels 5; Gaps			
Db	11	mltafm1vfum-afdsasaddysvveehqqlsisngeivelndrgepvqlgmshglqw	69
QY	7	IFAVLLMTALFSIGNTTAADYSVVEEHGOLSIISNGELVNERGEVOVLKGMSSHGLWY	66

Db 70 gqfvnyesmkwlrddwgitvfraamytssgyiedpsvkekveaeadlgiyviidw 129
 QY 67 GQFVNYESMKWLRDDWGITVFRAMYTSSGYIDPSVKEKVEAEADLGIYVIIDW 126
 Db 130 hilsdnpnylkeakdfdemselgdydpnvviyeianepngsdvtdnqikpyaevip 189
 QY 127 HILSDNDPNYIKEAKDFDEMSELYGDPNPVYIEIANEPNGSDVTWQNKPYAEVIP 186
 Db 190 virndpnnliivgtwsgdvhaadnqltdpnmvafhyagthgnlrdrqdyaldq 249
 QY 187 VIRNDPNNVIVGTWSDVHAADNQLADPNVAFHYAGTHGNLRDQDYALDQ 246
 Db 250 gaafvsewtsatgsgvfldeaqvdfmdernlswanlsthkdesaalmpgasp 309
 QY 247 GAAIFVSEWTSATGSGVFLDEAQVDFMDESLNLSWANSLTHKDESSAALMPGANP 306
 Db 310 tggwteaelspgfvfvekreiareattppsdptppsdpgpepdpptpsdpgdy 369
 QY 307 TGGWTEAELSPGFTVREKIREASIPSPDPTPPSDP--GEP--DPGEPDPTPPSDPGEY 362
 Db 370 pawdntiyltdelvyhngqlwqakwtqngqpgdpgypwepln 412
 QY 363 PAWDSNQIYNEIVYHNGQLWQAKWTQNGQPGDYPGWPEPLK 405

RESULT 12

ID R42122 standard; Protein; 499 AA.
 AC R42122;
 DT 27-APR-1994 (first entry)
 DE NK-1 cellulase.
 KW Cellulase; pH dependence; mutation.
 OS Bacillus N4.
 PN J05236959-A.
 PD 17-SEP-1993.
 PF 28-FEB-1992; 075883.
 PR 28-FEB-1992; JP-075883.
 PA (BEPF) BEPPU T.
 DR WPI; 93-330585/42.
 DR N-PSDB; Q49820.
 PT Changing the pH-dependence of cellulase enzymatic activity - by
 changing base sequence of cellulose-producing gene of *Bacillus*
 microbe to base sequence coding asparagine and serine residues at
 specified aminoacid sites
 S Disclosure: Fig 1: 9pp; Japanese.
 CC Sequences (Q49818-19) consist of two synthetic fragments which are
 used to induce a mutation within the cellulase NK-1 gene. The gene
 is shown in sequence (Q49820). The modified NK-1 gene shows a
 CC change in pH dependence.
 CC Sequence 499 AA;

Query Match 48.2%; Score 1649; DB 8; Length 499;

Best Local Similarity 62.7%; Pred. No. 4.80e-115;
Matches 222; Conservative 62; Mismatches 61; Indels 6; Gaps 6;

Db 1 mkrsisifitcllitvmtgqlqspasaaqtktpaakngqlskgtqlvnrdrkavqlk 60
 QY 1 MKKITTFIV-LMT-LALFSI-GNTAAADYSV-VEEHGQLSNGELVNERGEQVLK 56
 Db 61 gishgqlwqvgdvfnkdsllkwrddwgitvfraamytdaggyidnpvsnkveaveak 120
 QY 57 GMSHGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGGYIDPSVKEKVEAEVAAI 116
 Db 121 elgylviidwhilndgnpnghkekakoffkemsslyngtppnviyeianepng-dvnrkrd 179
 QY 117 DLGIYVIIDWHILSDNDPNYIKEAKDFDEMSELYGDPNPVYIEIANEPNGSDVTWQNK 176
 Db 180 ikpyaevisvirkndpndnliivgtgtwsgdvdaaddqlkdanvmyalhfyagthgsl 239
 QY 177 IKPYAEVIVIRNDPNNVIVGTGTWSDVDAADDQLKDPNVMYAFHYAGTHGQNL 236
 Db 240 rdkanyalskgaplfvtewtsdsgngvflqdsrewnlyldeksnlswnwnlsdkges 299

QY 237 RDQVDYALDQGAALFVSEWTSATGSGVFLDEAQVDFMDESLNLSWANSLTHKDES 296
 Db 300 ssalkpgasktgwpltdltasgtfvrenilgnkdskterpetpaqdnpageng 353
 QY 297 SAALMPGANPTGGWTEAELSPGFTVREKIREASIPSPD-TPPSDPGEPDPG 349

RESULT 13

ID W18790 standard; protein; 551 AA.
 AC W18790;
 DT 18-NOV-1997 (first entry)
 DE Corrected *Bacillus lautus* (NCIMB 40250) endoglucanase Endo 3A.
 KW Endoglucanase; Endo 3A; formation; localised; variation;
 KW colour density; surface; dye; fabric; family 5; cellulose;
 KW hydrolysatoin; p-nitrophenyl-beta-1,4-cellobioside; stone wash;
 KW blue jeans; back staining.
 OS *Bacillus lautus*.
 PN W09709410-A1.
 PD 13-MAR-1997.
 PF 03-SEP-1996; DK0364.
 PR 08-SEP-1995; DK-000993.
 PA (NOVO) NOVO-NORDISK AS.
 PI Fich M, Onishi M, Schulein M, Toft AH;
 DR WPI; 97-192888/17.
 PT Localised variation of colour density in the surface of a dyed
 cellulose fabric - uses cellulase compsn. able to hydrolyse
 p-nitrophenyl-beta-1,4-cellobioside
 PS Disclosure: Pages 15-17; 23pp; English.
 CC The present sequence is the corrected version of the incorrect
 CC *Bacillus lautus* (NCIMB 40250) endoglucanase Endo 3A described in
 CC W09110732. Endo 3 can be used in novel method of forming localised
 CC colour density variation on the surface of a dyed cellulosic
 CC fabric. The method comprises agitating the fabric in an aqueous
 CC medium (pH 6.5 to 9.0) containing a family 5 cellulose,
 CC e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta
 CC -1,4-cellobioside, or a family 7 cellulase, and a mechanical
 CC abrading agent or cellulose having abrading activity. Each
 CC cellulase displays 30 % or more of its maximum activity at pH 7.
 CC The process is useful to provide a stone washed look to blue jeans
 CC without back staining.
 CC Sequence 551 AA;

Query Match 45.7%; Score 1561; DB 24; Length 551;

Best Local Similarity 62.3%; Pred. No. 3.38e-108;
Matches 192; Conservative 63; Mismatches 51; Indels 2; Gaps 2;

Db 7 gqkvqngqlvgqsgavqlvgmshglqwygnfnvkslqwmrdnwglnvfraamytae 66
 QY 36 GQLSINGELVNERGEQVLKGMSSHGLQWYGFVNYESMKWLRDDWGITVFRAMYTSS 95
 Db 67 dgyitdpsvknkvkeavqasidlglyviidwhilndgnpnntykaqskaffgematlyngt 126
 QY 96 GGYIDDPSSVKEKVEAEADLGIYVIIDWHILSDNDPNYIKEAKDFDEMSELYGDP 155
 Db 127 pnviyeianepng-nvswad-vksyaeveitairaidpdpvgvipsptsqdihaadnp 184
 QY 156 PNVIYEIANEPNGSDVTWQNKPYAEVIVIRNDPNNVIVGTGTWSDVHAADNQ 215
 Db 185 vshsnvmyalhfyagthgqflrdityamnkgaalfvtewtsdsgngpgyfpqskewi 244
 QY 216 LADPNVMTAFHYAGTHGNLRDQDYALDQGAALFVSEWTSATGSGVFLDEAQVWI 275
 Db 245 dflnarkiswvnwsladkvetsaalmpgastgvgvtdaqlesgkwvrdqirqtggsg 304
 QY 276 DFMDERNLSWANSLTHKDESSAALMPGANPTGGWTEAELSPGFTVREKIREASIPSP 335
 Db 305 nptapaap 312
 QY 336 DPTPPSDP 343

RESULT 14

ID W01503 standard; protein; 531 AA.

AC W01503;
 AT 26-FEB-1997 (first entry)
 DE 60 kD endoglucanase, EG C.
 KW Detergent composition; cellulase; retaining-type activity; catalytic;
 KW activity; cellotriose; particulate soil removal; colour clarification;
 KW cleaning; cellulose-containing fabric; cellubiohydrolase; endoglucanase.
 OS Bacillus lautus, NCIMB 40250.
 PN W09502675-A1.
 PD 26-JAN-1999;
 PE 07-JUL-1994; DK0280.
 PF 12-JUL-1993; EP-870131.
 PR 11-OCT-1993; DK-001135.
 PA (NOVO) NOVO-NORDISK AS.
 PA (PROC) PROCTER & GAMBLE CO.
 PI Convents AC, Jeffreys B, Schuelein M, Tikhomirov DF;
 DR WPI; 95-067325/09.
 PT Detergent compsn. contg. two cellulase components - the first
 PT removing soil particles and the second capable of colour
 PT clarification, useful in laundry compsns.
 PS Claim 28; Page 68-70; 83pp; English.
 CC Detergent compositions comprising: 1) a first cellulase component having
 CC retaining-type activity, pref. having a catalytic activity on cellotriose
 CC at pH 8.5 corresponding to that of at least 0.01 s<-1>, and capable of
 CC particulate soil removal; and 2) a second cellulase component having
 CC multiple domains comprising at least one non-catalytic domain attached to
 CC a catalytic domain, pref. having a catalytic activity on Red Avicel 7.5
 CC per 1 mg of cellulase protein higher than 10<-4> IU and being capable of
 CC colour clarification, where at least one of the cellulase components is a
 CC single (recombinant) component, are useful for cleaning and colour
 CC clarification of cellulose-containing fabrics. The second cellulase
 CC component can be an endoglucanase which is immunoreactive with an
 CC antibody raised against a highly purified ~60 kD endoglucanase
 CC derived from Bacillus lautus, NCIMB 40250, and is esp. the present
 CC sequence, designated EG C.
 SQ Sequence 531 AA;

Query Match 22-28; Score 1339; DB 20; Length 531;
 Best Local Similarity 58.5%; Pred. No. 5.84e-91;
 Matches 176; Conservative 52; Mismatches 68; Indels 5; Gaps 4;
 Db 13 vltmvlmgl1llpygarkgyaa-pavp-fgelkvqgnqlvgsgqavqlvgmsshglqwy 72
 QY 10 VLMTLALFSGNTTAAADDYSVVEEHGQLSISNGELVNERGEQVQLKGMSSHGLQWY 68
 Db 73 fvnksslqwmrdnwginvfraamytsgdyitdpsvknkveavqasidlalyviidwhi 132
 QY 69 FVNYESMKWLRDDMGITVFRAMYTSSGGYIDDPVKEKVEAEADLGIYVIIDWHI 128
 Db 133 lsdgnpntykaqskaffematiyngtnpviyeiarstec-vl--grcqs-seevitai 188
 QY 129 LSONDPNIYKEAKDFDEENSELYGDPNVIYEIANEPNGSDVTWQNDQIKPYAEVIVP 188
 Db 189 rlsidpdgsvivgspwtsqdihiadnpvshnvmvalhfygsthgqfdrityamnkga 248
 QY 189 RNDPNPNIYVIGTWSQDVHHAADNPVNYAFHFYAGTHGQNLRDQVYALDQGA 248
 Db 249 aifvtewgtsdasnggpylpqskewidflnarkiswvnwsladkvettsaalmgasp 308
 QY 249 AIFVSEWGTSAATGCGVFLDEAQVWIDFMDERNLSWANNSLTHKDESSAALMPCANFTG 308
 Db 309 a 309
 QY 309 G 309

RESULT 15
 ID R13229 standard; Protein; 532 AA.
 AC R13229;
 DT 14-OCT-1991 (first entry)
 DE Endoglucanase encoded by endo3 gene.
 KW Cellulase activity; detergent.
 OS Bacillus spp. NCIMB 40250.
 FH Key Location/Qualifiers

FT cleavage_site 36..37
 FT peptide 1..36
 FT /label= signal peptide
 FT peptide 37..531
 FT /label= mature peptide
 PN W09110732-A.
 PD 25-JUL-1991.
 PE 18-JAN-1991; DK0013.
 PF 19-JAN-1990; DK-000164.
 PA (NOVO) NOVO NORDISK A/S.
 PI Jorgensen PL, Schulein M, Hansen C;
 DR WPI; 91-238020/32.
 DR N-PSDB; Q13003.
 PT Enzyme exhibiting cellulase activity from Bacillus sp. - is an
 PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
 PT fabrics.
 PS Claim 1; Page 80; 96pp; English.
 CC The enzyme is encoded by a 11000 bp. EcoRI fragment of Bacillus
 CC spp. R1236 DNA contained in plasmid pPL591. It exhibits an
 CC endoglucanase activity of at least 10 (pref. at least 25)
 CC carboxymethyl cellulose (CMC) endoase units per mg total protein
 CC under alkaline conditions. It is especially useful as a
 CC cellulytic agent and has been found to be more stable during
 CC washing (60 mins. at 40 deg.) in the presence of conventional
 CC detergents than a commercial cellulase preparation. It may also
 CC show increased storage stability in liq. detergents contg.
 CC proteases. The sequence was deduced from the DNA (Q13003), it is a
 CC product of the endo3 gene. See also R13227 and R13228.
 SQ Sequence 532 AA;

Query Match 38-98; Score 1329; DB 3; Length 532;
 Best Local Similarity 56.8%; Pred. No. 3.48e-90;
 Matches 172; Conservative 61; Mismatches 64; Indels 6; Gaps 5;
 Db 13 vltmvlmgl1llpygarkgyaa-pavp-fgelkvqgnqlvgsgqavqlvgmsshglqwy 70
 QY 7 fVAVLMTLALFSGNTTAAADDYSVVEEHGQLSISNGELVNERGEQVQLKGMSSHGLQWY 66
 Db 71 gnfvnksslqwmrdnwginvfraamytsgdyitdpsvknkveavqasmdlalyviidw 130
 QY 67 GQFVNYESMKWLRDDMGITVFRAMYTSSGGYIDDPVKEKVEAEADLGIYVIIDW 126
 Db 131 hilsdgnpntykaqskaffematiyngtnpviyeiarstec-vl--grcqs-seevit 186
 QY 127 HILSDNDPNIYKEAKDFDEENSELYGDPNVIYEIANEPNGSDVTWQNDQIKPYAEVIVP 186
 Db 187 airdsidpdgsvivgspwtsqdihiadnpvshnvmvalhfygsthgqfdrityamnk 246
 QY 187 VIRNDPNPNIYVIGTWSQDVHHAADNPVNYAFHFYAGTHGQNLRDQVYALDQ 246
 Db 247 gaafvtewgtsdasnggpylpqskewidflnarkiswvnwsladkvettsaalmgasp 306
 QY 247 GAAPVSEWGTSAATGCGVFLDEAQVWIDFMDERNLSWANNSLTHKDESSAALMPCANP 306
 Db 307 tga 309
 QY 307 TGG 309

Search completed: Fri Jun 4 10:00:43 1999
 Job time : 82 secs.

MPERF

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Search_pp protein - protein database search, using Smith-Waterman algorithm
on: Fri Jun 4 10:01:50 1999; MasPar time 34.60 Seconds
905,518 Million cell updates/sec
Tabular output not generated.

Title: >US-08-945-574-2
Description: (1-574) from US08945574.pep
Perfect Score: 4235
Sequence: 1 MKWKMVMVLAVLVSFVA.....GNQVTGIAAQTTSKNKKK 574

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sp_trembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_invertebrate 14:sp_virus

Statistics: Mean 50.877; Variance 94.403; scale 0.539

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	604	14.3	414	2	BETA-1,4-D-GLUCANASE (1.25e-97
2	567	13.4	477	3	CELLULOSE	1.08e-89
3	535	12.6	400	2	CELLULOSE (EC 3.2.1.4)	7.24e-83
4	535	12.6	1232	3	CELLULOSE CELD (FRAGME	2.23e-81
5	528	12.5	482	3	CELLULOSE (FRAGMENT)	4.81e-79
6	517	12.2	471	3	CELLULOSE	3.38e-78
7	513	12.1	388	3	CELLULOSE	3.92e-69
8	470	11.1	455	2	ENDOGLUCANASE A PRECUR	4.38e-68
9	465	11.0	759	2	CARBOXYMETHYLCELLULOSE	2.53e-64
10	447	10.6	363	2	B14, PUTATIVE POLYGALA	9.06e-54
11	447	10.6	924	2	ENDOGLUCANASE CELG.	2.99e-35
12	396	9.4	519	2	XYLANASE PRECURSOR.	3.51e-08
13	304	7.2	584	2	ENDOGLUCANASE 2 (EC 3.	8.25e-06
14	156	3.7	490	2	ENDOGLUCANASE 2 (EC 3.	3.79e-04
15	141	3.3	419	3	ENDOGLUCANASE 2 (EC 3.	1.46e-03
16	137	3.2	410	3	ENDOGLUCANASE 2 (EC 3.	3.94e-03
17	130	3.1	416	3	ENDOGLUCANASE 2 (EC 3.	1.45e-02
18	126	3.0	368	3	ENDOGLUCANASE 2 (EC 3.	
19	123	2.9	494	2	ENDOGLUCANASE 2 (EC 3.	
20	119	2.8	160	3	EXO-1,3-BETA-GLUCANASE	

21	113	2.7	214	5	Q18547	9.71e-02
22	115	2.7	430	3	Q12712	5.19e-02
23	113	2.7	597	2	Q51037	9.71e-02
24	109	2.6	144	2	P78000	3.32e-01
25	112	2.6	363	2	Q07652	1.32e-01
26	110	2.6	481	2	Q68652	2.45e-01
27	109	2.6	516	2	Q84333	3.32e-01
28	109	2.6	570	2	Q59665	3.32e-01
29	110	2.6	744	2	Q06693	2.45e-01
30	109	2.6	1335	14	Q89353	3.32e-01
31	107	2.5	173	2	Q51892	6.07e-01
32	105	2.5	344	3	Q14405	1.10e+00
33	104	2.5	400	2	Q85465	1.10e+00
34	105	2.5	468	5	Q17761	1.10e+00
35	104	2.5	516	1	Q50221	1.47e+00
36	105	2.5	567	1	Q38628	1.10e+00
37	107	2.5	607	3	Q01421	6.07e-01
38	106	2.5	608	4	Q60378	8.18e-01
39	107	2.5	1802	10	Q48647	6.07e-01
40	102	2.4	299	4	Q92984	2.63e+00
41	103	2.4	361	2	Q60100	1.97e+00
42	102	2.4	455	2	Q33340	2.63e+00
43	102	2.4	872	13	Q57392	2.63e+00
44	103	2.4	1257	11	Q54943	1.97e+00
45	103	2.4	1257	11	Q54954	1.97e+00

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	414 AA.
ID Q59733			
AC Q59733;			
DT 01-NOV-1996 (TREMUREL. 01, CREATED)			
DT 01-NOV-1996 (TREMUREL. 01, LAST SEQUENCE UPDATE)			
DT 01-NOV-1996 (TREMUREL. 08, LAST ANNOTATION UPDATE)			
DE BETA-1,4-D-GLUCANASE (EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)			
DE (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).			
GN CELA.			
OS RUMINOCOCCUS ALBUS.			
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;			
OC RUMINOCOCCUS.			
RN [1]			
RP SEQUENCE FROM N.A.			
RA VERCOR P.E., GREGG K.;			
RL SUBMITTED (FEB-1993) TO EMBL/GENBANK/DBJ DATA BANKS.			
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC			
CC LINKAGES IN CELLULOSE.			
DR EMBL; L10243; G152631; -			
DR PFAM; PF00150; cellulase; 1.			
KW HYDROLASE; GLYCOSIDASE.			
SQ SEQUENCE 414 AA; 45879 MW; 42453336 CRC32;			

Query Match	14.3%;	Score 604;	DB 2;	Length 414;
Best Local Similarity	35.8%;	Pred. No. 1.25e-97;		
Matches	101;	Conservative	67;	Mismatches 96; Indels 18; Gaps 13;
Db	70	EVRDISAMELVAMKGTGNLNSLSDATGAAGNASEVNGNPKTKEMIDAVYKGFVIR	129	
Qy	29	DKVTLDIQSYVRDQPGWNLGNTDAVGQ--D--ETANGNPRVTRIELIERIADGKYSIR	84	
Db	130	IPVTWGGHVGDPYKIDENLWLRVQEVVYAYDGAIVINSHEELURIPDNHI-DA	188	
Qy	85	IPVTWENRIGGAPDYPIDPQFLNRYVDEYVVALEEDLYVIMLNHSDSLWIYEMEHNYG	144	
Db	189	VDETAALWQVAERFKDYGDLIFEGNEPRVKSGPEWNGTTEGRRCV-ERLNQVFL	247	
Qy	145	VMAVRSLEWQENHFKDYPKLMFESVNEPFF--S-QNW-GEIRENHALLDDLTVFF	200	
Db	248	DTVRATGNNKRR-LLL--MTTYASSCGLK-LIQDTAIPEDDDHIGFSHAYTPYAFYNA	303	
Qy	201	EIVRSGSGQNDIRPLVLTMTATSQPLNLLNYQIDKLDPNLATVHYGFWFFSVNI	260	
Db	304	NADWELFENDWNPATAEILITLSNLUKENYLDKQDIPVITEYG	345	


```
RESULT 5
ID Q01409 PRELIMINARY; PRT; 482 AA.
AC Q01409;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE CELLULASE (FRAGMENT).
GN CELA.
OS NEOCALLIMASTIX FRONTALIS (RUMEN FUNGUS).
OC EUKARYOTA; FUNGI; CHYTRIDIOMYCOTA; CHYTRIDIOMYCETES;
OC NEOCALLIMASTICALES; NEOCALLIMASTICACEAE; NEOCALLIMASTIX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MCH3;
RX MEDLINE: 99022200.
RA FUJINO Y., OGATA K., NAGAMINE T., USHIDA K.;
BT "Cloning, sequencing, and expression of an endoglucanase gene from
the rumen anaerobic fungus Neocallimastix frontalis MCH3.";
BIOSCI. BIOTECHNOL. BIOCHEM. 62:1795-1798(1998).
DR EMBL: J38843; G3712668; -.
FT NON-TER 1
SQ SEQUENCE 482 AA; 54646 MW; B5815F31 CRC32;

Query Match 12.5%; Score 528; DB 3; Length 482;
Best Local Similarity 30.8%; Pred. No. 2.23e-81;
Matches 88; Conservative 81; Mismatches 99; Indels 18; Gaps 15;

Db 78 ETCWGNPKTDMFKVMDNQNFVRIPTTWSGHFGADPYKINEKWLKRVHEIVDYPK 137
QY 59 ETAWGNPRVTRELIERIADEGYKSIRIPVTWENRIGGAPYDIPQFLNRVDEVVQWALE 118

Db 138 NGAFVILNHHTWNAHFS-E-TLDTAKEILEKIQSIAKEFKDYDDEHLIFGGLNEPRKN 195
QY 119 EDLYVMINLHDSWLWIYEMEHNYNGVMKYSRLWEQLSNHFKDYPYTKLMFESVNEP-KF 177

Db 196 DTPVEVTGGQEGWDV-NAMNAVLKTIIRSSGNNPKRHLMIPTPYAAACNENSKN-F- 252
QY 178 -SQ-NW-GEIRENHALLDDLTNTVFEIVRQSGQNDIRPLVPTMETATSQPLNNLYQ 234

Db 253 IPPE-DDDKVIAVHAYAPYNFALNGAGAVDKFADAGKKDLWNINLKKRFFVDQIGPM 311
QY 235 TIDKLDPNLIATVHYIGFWPFSVNI-AGYT-RFEEDSKREIETFDVRVHHTFVARGIPV 292

Db 312 ILGEYGM--NRDN---EE-ERATWAEFYMEKYVTAMGVQVQWMDNG 351
QY 293 VLGEFGLGDKHTGVIQQGEKLFEEYLIHHLNERDITHLMDNG 338

RESULT 6
ID P78719 PRELIMINARY; PRT; 471 AA.
AC P78719;
DT 01-WAY-1997 (TREMREL. 03, CREATED)
DT 01-WAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE CELLULASE.
GN CELB.
OS ORPINOMYCES SP.
OC EUKARYOTA; FUNGI; CHYTRIDIOMYCOTA; CHYTRIDIOMYCETES;
OC NEOCALLIMASTICALES; NEOCALLIMASTICACEAE; ORPINOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PC-2;
RX MEDLINE: 97176394.
RA LI X.L., CHEN H., LJUNGDAHL L.G.;
RT "Monocentric and polycentric anaerobic fungi produce structurally
related cellulases and xylanases."
RL APPL. ENVIRON. MICROBIOL. 63:628-635(1997).
DR EMBL: U57818; G1688087; -.
DR PROSITE: PS01159; WWDOMAIN_1; 1.
DR PFAM: PF00150; cellulase; 1.
SQ SEQUENCE 471 AA; 53103 MW; EA9C65EA CRC32;

Query Match 12.2%; Score 517; DB 3; Length 471;
```

```
Best Local Similarity 29.7%; Pred. No. 4.81e-79;
Matches 104; Conservative 99; Mismatches 123; Indels 24; Gaps 19;

Db 1 MKFLNSLSLGLVIAGCEAMRNIS-KELVKELTIGNSGLTLDASCVETLNYSKDQTAS 59
QY 1 MKWKMSVWLAVLVVSFVAPAVSSANEDVKTLDIQ-SYVRDMQPGWNLNTFD-AVQOD 58

Db 60 ETCWGNVKTQELYYKLSDLGNTFRIPPTWSGHFGADPYKISDVMMKRVHEVVVDYALN 119
QY 59 ETAWGNPRVTRELIERIADEGYKSIRIPVTWENRIGGAPYDIPQFLNRVDEVVQWALE 118

Db 120 TGGYAILNIHETW--NYAFQKNLESAKKILVAIKQIAAEFGDYDEHLIFEGMNEPKV 177
QY 119 EDLYVMINLHDSWLWIYEMEHNYNGVMKYSRLWEQLSNHFKDYPYTKLMFESVNEP-KF 177

Db 178 GDPAEVTGGQEGWN-FVNMENALFVKTIIRATGNNRNLHMIPTPYAAASVNDGSINN-FK 235
QY 178 SQ--NW-GEIRENHALLDDLTNTVFEIVRQSGQNDIRPLVPTMETATSQPLNNLYQ 234

Db 236 YPNG-DD-KVIVSLHSYSPYNFALNNGPGAISNFY-DGN-EIDWYMTNTINSFISKGIPV 291
QY 235 TIDKLDPNLIATVHYIGFWPFSVNI-A-G-YTREEDSKREIETFDVRVHHTFVARGIPV 292

Db 292 IIGEF--VAMNRDN---ED-DRERQWQYIYKATAGIPCVIWDNG-YFE 334
QY 293 VLGEFGLGDKHTGVIQQGEKLFEEYLIHHLNERDITHLMDNGQHFN 342

RESULT 7
ID Q13334 PRELIMINARY; PRT; 388 AA.
AC Q13334;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE CELLULASE.
GN CELB2.
OS ORPINOMYCES JOYONII.
OC EUKARYOTA; FUNGI; CHYTRIDIOMYCOTA; CHYTRIDIOMYCETES;
OC NEOCALLIMASTICALES; NEOCALLIMASTICACEAE; ORPINOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SG4;
RA QIU X., SELINGER L.B., YANKE L.J., CHENG K.-J.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF015249; G2353007; -.
DR PFAM: PF00150; cellulase; 1.
SQ SEQUENCE 388 AA; 43835 MW; 1DBEBD73 CRC32;

Query Match 12.1%; Score 513; DB 3; Length 388;
Best Local Similarity 31.1%; Pred. No. 3.38e-78;
Matches 89; Conservative 78; Mismatches 101; Indels 18; Gaps 15;

Db 29 ETCWGNPKTDMFKVMDNQNFVRIPTTWSGHFGADPYKIDKMKRVHEVVVDYAYN 88
QY 59 ETAWGNPRVTRELIERIADEGYKSIRIPVTWENRIGGAPYDIPQFLNRVDEVVQWALE 118

Db 89 NGAFVILNHHTWNAHFS-E-TLETAKVILADTNQIAAEFEKYNERKLIFEGLEPRKN 146
QY 119 EDLYVMINLHDSWLWIYEMEHNYNGVMKYSRLWEQLSNHFKDYPYTKLMFESVNEP-KF 177

Db 147 DTPVEWNGDGKEGWDV-NAMNEVFLKTIIRASGNNPKRHLMIPTPYAAATQENSEKN-FK 204
QY 178 -SQ-NW-GEIRENHALLDDLTNTVFEIVRQSGQNDIRPLVPTMETATSQPLNNLYQ 234

Db 205 YPGG-DD-KLVYSVHNYAPYNFALNNGDGAVETDAQCKDLDWSISLIKTFDTKGPV 262
QY 235 TIDKLDPNLIATVHYIGFWPFSVNI-A-GYT-RFEEDSKREIETFDVRVHHTFVARGIPV 292

Db 263 IMGEYGMNRD-NT-----E-BRAKWAEEYMEKYVTAGVQVQWMDNG 302
QY 293 VLGEFGLGDKHTGVIQQGEKLFEEYLIHHLNERDITHLMDNG 338
```

DR	PFAM; PF00404; celCC; 1.
KW	SIGNAL.
FT	1 33 POTENTIAL.
FT	34 759 ENDUGLUCANASE A.
QY	SEQUENCE 759 AA; 83813 MW; F21E25E3 CRC32;
Query Match	11.0%; Score 465; DB 2; Length 759;
Best Local Similarity	29.4%; Pred. No. 4.38e-68;
Matches	78; Conservative 67; Mismatches 103; Indels 17; Gaps 16
Dd	75 LDSAGLETETCWGCPASQELFADKAAGKGTNVRPTTFQHLDEN--DN-IDPAWMARVH 132 ::: : - :: : :: : :
Qy	52 FDVGQD-FATWGPNRVTRIELIERIADEGYKSIRIPVTWENRIGGAPDYDPQLNRVD 110 ::: : : : : : : :
Dd	133 QVVDYAYNGIYVINLHHEQNINRRDLATAVDDINPLMKLTQTATFEKYDQHLIF 192 ::: : : : : : : :
Qy	111 EVQWALEEDLYVMNLHHD-SLWIYEMEHNTGVNAKYSRWEQSUNHFKOYPTKLMF 169 ::: : : : : : : :
Dd	193 ECMNEPRAMDTPEWSWATSATPVEADYVLRLEAFVELIRMGIDGPYAKTRLMLPGKVASS 252 : : : : : : : : : : :
Qy	170 ESNVEPK-F-SQ-NW-GEIRENHALLDDLNTVFVEIVRQSGGQ-NDIRPLVLPTMETAT 224 ::: : : : : : : : : : :
Dd	253 DKTELNOIVLP-EN-DD-FIATSIHAYTPNFMTNKTEBGAHVDTFKESFNDAYNLQ 309 ::: : : : : : : : : :
Qy	225 SQPLLNNLYOTDKLDDPNLIATVHYGYGFWPFSVNI-A-G-Y-TRFEEDSKREIIETF 279 ::: : : : : : : : : :
Dd	310 NFRDMFINKDIPPVIGMGTSDRCN 334
Qy	280 RVHTFTVARGIPVVIGEGFLGFDK 304 : : : : : : : : : :
RESULT	10
ID	Q44878 PRELIMINARY; PRT; 363 AA.
AC	Q44878
DT	01-NOV-1996 (TREMBLREL_01, CREATED)
DT	01-NOV-1996 (TREMBLREL_01, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBLREL_08, LAST ANNOTATION UPDATE)
DE	CARBOXYMETHYLCELLULOSE.
GN	CWCASE.
OS	BACTEROIDES RUMINICOLA.
OS	BACTERIA; CYTOPHAGALES; BACTEROIDACEAE; PREVOTELLA.
OC	[1]
RN	SEQUENCE FROM N.A.
RP	MEDLINE; 90299778.
RX	MATSUSHITA O., RUSSELL J.B., WILSON D.B.;
RA	"Cloning and sequencing of a Bacteroides ruminicola B(1)4
RT	endoglucanase gene."
RL	J. BACTERIOL. 172:3620-3630(1990).
RL	EMBL; M38216; GI43941; "
DR	PFAM; PF00150; cellulase; 1.
QY	SEQUENCE 363 AA; 40526 MW; 4877CE66 CRC32;
Query Match	10.6%; Score 447; DB 2; Length 363;
Best Local Similarity	29.2%; Pred. No. 2.53e-64;
Matches	79; Conservative 69; Mismatches 108; Indels 15; Gaps 11:
Dd	50 PVATYETFWGPPETQDMTFFLMQNGFNAYRIPIVTVYEHMDAEGN--VDEAWMRVAIV 107 ::: : : : : : : : : :
Qy	54 AVGODETAWNPNRVTRIELIERIADEGYKSIRIPVTWENRIGGAPDYDPQLNRVDEV 113 ::: : : : : : : : : :
Dd	108 EYAMNAGLYAVNVVHHDTAAAGSAWTKADTDVYAAYKEKFKKWTOTIANALADYDQHLIF 167 : : : : : : : : : : : : :
Qy	114 QWALEEDLYVMNLHDS---W-LWIYEMEHNTGVNAKYSRWEQSUNHFKOYPTKLMF 169 ::: : : : : : : : : :
Dd	168 EGYNEMLDGNSWNDEPOKASGYPALNNYAODFYDVAVRATGGNNATRNLIYNTYAAAAGEN 227 : : : : : : : : : : : : :
Qy	170 ESNVEP-KFSQNWGEI-RENHALLDDLNTVFVEIVRQSGGQNDIRPLVLPTMETATSQ 227 ::: : : : : : : : : :
Dd	228 VLNFMFLPTDAVN-HLIVOVHSYDPWFN-FNKT-TW-DSECHNLTIEIFSALSCKFTT 293 ::: : : : : : : : : :
Qy	228 LLNNLYOTDKLDDPNLIATVHYGYGFWPFSVNIAGYTRFEEDSKREIIETFDRVHHTFA 287 ::: : : : : : : : : :
Dd	284 --IPYIGEYTHGESDISVSKSSPAFKIKL 312

DB 13 VAFAGSVSAATLPTAK-EVQA--K-MGMFNIGNSME-VPNSTLNGNPPYTPQPLDLSV 67
::: ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 16 VSFAFPAVSANEDVKTLDIQSYYVRDMQPGWNLGNTDFAVGDQDETANGPRVTRELIERI 75
::: ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Db 68 KAAGENTVRIPCAWDSHSGS-K-VTET-WLDSVKTVDVAMRAGLTILNIHHEGGGW 124
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 ADGYKSIRIPVTWENRIGGAPDPIDPQLNRVDEVVQWALEEDLYVMINLHDS-WLW 134
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 125 FQSNIGTSVDTNDKNMKTWTQTIANFKDYNERLLFAGANEPPNVNTW--TSQHVTOL 182
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 I-YEMEHNYGVMA-KYRSLEWELSLNFHKDYPTKLMEFSYNPFQSQ-NWGEIRENHIAL 191
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 183 MHIYT-FIDAVRATGNNATRLLIIQGLNTRIDDKSVKSPVTFPKDKVGRLMFEVHY 241
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 LDDLNTVFEEIVRQSGQNDIRPLVLPMTMETATSQPLLNLYQITDKLD-DPNLIATVRY 250
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 242 YDYPQYTL 249
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 251 YGFWPFSV 258
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
ID Q45397 PRELIMINARY; PRT; 584 AA.
AC Q45397;
DT 01-NOV-1996 (TREMBREL. 01, CREATED)
DT 01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE XYLANASE PRECURSOR
OS BACTEROIDES RUMINICOLA
OC BACTERIA; CYTOPHAGALES; BACTEROIDACEAE; PREVOTELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23;
RX MEDLINE; 93291665.
RA WHITEHEAD T.R.;

RT "Analyses of the gene and amino acid sequence of the Prevotella
RT (Bacteroides) ruminicola 23 xylanase reveals unexpected homology with
RT endoglucanases from other genera of bacteria."
RL CURR. MICROBIOL. 27:27-33(1993).
DR EMBL; M83379; G143974; -;
DR PFAM; PF00150; cellulase; 1.
KW SIGNAL; XYLAN DEGRADATION; HYDROLASE; GLYCOSIDASE.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 584 XYLANASE.
SQ SEQUENCE 584 AA; 65740 MW; 7272FA75 CRC32;

Query Match 7.2%; Score 304; DB 2; Length 584;
Best Local Similarity 38.4%; Pred. No. 2.99e-35;
Matches 48; Conservative 31; Mismatches 37; Indels 9; Gaps

Db 67 ESETCQGSKAKPELIKMKMDAGFGAIRPVTVWNHMDKGK--VNDEMRRVHEVDYV 124
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 57 QDETANGNPVRTIELIERIADGYKSIRIPVTWENRIGGAPDPIDPQLNRVDEVQWA 116
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 125 INOGLYCVNVHHDGTGSDSFKSWIKADEANYTNKRIAYENLWQIAEERFDYDEHLFF 184
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 LEEDLYVMINLHD----S-WL--WIYEMEHNYGVAKYSLWEQLSNHFKDPTKLMF 169
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 185 ESYNE 189
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 ESYNE 174
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
ID Q33853 PRELIMINARY; PRT; 490 AA.
AC Q33853;
DT 01-JAN-1998 (TREMBREL. 05, CREATED)
DT 01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE ENDOGLYCOCERAMIDASE II (EC 3.2.1.123).
OS RHODOCOCCUS SP.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERIAE; NOCARDIACEAE; RHODOCOCCUS.

[illegible]


```
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-M-777;
RX MEDLINE; 97390413.
RA IZU H., IZUMI Y., KURUME Y., SANO M., KONDO A., KATO I., ITO M.;
RT "Molecular cloning, expression, and sequence analysis of the
RT endoglycoceramidase II gene from Rhodococcus species strain M-777.";
RL J. BIOL. CHEM. 272:19846-19850(1997).
DR ENBL; U39554; G2337906; -.
KW HYDROLASE; GLYCOSIDASE.
SQ SEQUENCE 490 AA; 52750 MW; F4E87E0A CRC32;

Query Match      3.7%; Score 156; DB 2; Length 490;
Best Local Similarity 29.8%; Pred. No. 3.51e-08;
Matches 34; Conservative 31; Mismatches 40; Indels 9; Gaps 9;

Db 29 AASGSGSGTAL-TPSYLKD-DGGRSILRGFNFTASSAKSAPDGMPOFTADLAREYAD 86
   :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
QY 21 PAVSSANEDVKTLDIQSVYRDMPQGNLG-NTFDVAGQDETAW-CNPRVTR-ELIERIAD 77

Db 87 MGTNFRFLISWRS-VEPAPG-VYDQQYLDVRVEDRVGWYAERG-YKVMMDMHQD 137
   :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
QY 78 EGYKSIRIPVTWENRIGGAPDYPIDPQFLNRVDEVVQWALEEDLY-VMINLHHD 130

RESULT 15
ID Q12539 PRELIMINARY; PRT; 419 AA.
AC Q12539;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE EXO-1,3-BETA-GLUCANASE PRECURSOR (EC 3.2.1.58)
DE (GLUCAN 1,3-BETA-GLUCOSIDASE) (EXO-1,3-BETA-GLUCOSIDASE).
GN ABEXG1.
OS AGARICUS BISPORUS (COMMON MUSHROOM).
OC EUKARYOTA; FUNGI; BASIDIOMYCOTA; HYMENOMYCETES; AGARICALES;
OC AGARICACEAE; AGARICUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HORST U1;
RX MEDLINE; 96304295.
RA VAN DE RHEE M.D., MENDES O., WERTEN M.W.T., HUIZING H.J.,
RA MOOIBROEK H.;
RT "Highly efficient homologous integration via tandem exo-beta-1,
RT 3-glucanase genes in the common mushroom, Agaricus bisporus.";
RL CURR. GENET. 30:166-173(1996).
DR ENBL; X92961; G1064880; -.
KW SIGNAL; HYDROLASE; GLYCOSIDASE.
FT SIGNAL 1 22 POTENTIAL.
SQ SEQUENCE 419 AA; 46682 MW; 05344CCA CRC32;

Query Match      3.3%; Score 141; DB 3; Length 419;
Best Local Similarity 32.4%; Pred. No. 8.25e-06;
Matches 22; Conservative 19; Mismatches 23; Indels 4; Gaps 4;

Db 85 W-NTWITEEDFARIAAGLNHRLPIGYWAFEAAGEPY-IQGG-LPFLERKAVTWAQHN 141
   :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
QY 62 WGNPRVTRIELIERIADEGYSIRIPVT-WENRIGGAPDYPIDPQFLNRVDEVVQWALEED 120

Db 142 LKLIIDLH 149
   :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
QY 121 LYVMINLH 128
```

Search completed: Fri Jun 4 10:03:56 1999
Job time : 126 secs.

MPARLH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

on: Fri Jun 4 10:05:47 1999; MasPar time 22.00 Seconds

Tabular output not generated. 554.952 Million cell updates/sec

Title: >US-08-945-574-2

Description: (1-574) from US08945574.pep

Perfect Score: 4235

Sequence: 1 MKWKMVMYLAIVLVYSEFA.....GNQVTGIAQTNSRKNKK 574

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.906; Variance 180.757; scale 0.204

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	4235	100.0	574	19	W00383	Bacillus cellulase BC
2	2557	60.4	566	3	R13228	Endoglucanase encoded
3	597	14.1	406	2	R08199	Neutral cellulase gen
4	535	12.6	452	9	R49102	Translated sequence o
5	531	12.5	800	9	R47496	Translated sequence o
6	517	12.2	471	32	W56742	Orpinomyces cellulase
7	335	7.9	360	26	W34566	Thermotoga OCI/4V end
8	326	7.7	360	36	W49870	Thermotoga OCI/4V end
9	155	3.7	461	23	W10210	Mature endoglucoceram
10	156	3.7	490	23	W10209	Full length endoglyco
11	150	3.5	418	32	W57421	Amino acid sequence o
12	141	3.3	418	14	R79540	Endoglucanase-II prot
13	141	3.3	418	14	R77264	T. longibrachiatum en
14	141	3.3	418	18	W02032	Trichoderma endogluc
15	119	2.8	429	17	R88407	Beta-(1,6)-endoglucan
16	113	2.7	454	26	W34559	Thermococcus AEDII12R

17	113	2.7	454	36	W49863	Thermococcus AEDII12R
18	106	2.5	269	31	W28052	Amino acid sequence o
19	104	2.5	400	25	W22521	Bacillus agaradherens
20	104	2.5	400	25	W23601	Bacillus agaradherens
21	104	2.5	400	32	W57431	Bacillus agaradherens
22	104	2.5	462	32	W57433	Cloned alkaline endog
23	107	2.5	510	8	R47456	Truncated xylanase (X
24	107	2.5	551	8	R47455	Truncated xylanase (X
25	107	2.5	607	8	R47454	Xylanase (XYLA).
26	107	2.5	644	9	R44529	PNPX30 xylanase.
27	106	2.5	645	34	W29653	Human secreted protei
28	106	2.5	1969	36	W2419	Rice bacterial leaf b
29	100	2.4	176	19	R38153	Thermostable inorgani
30	100	2.4	272	36	W69432	Human secreted protei
31	103	2.4	551	24	W18790	Corrected Bacillus la
32	98	2.3	249	37	W2670	Canine herpes virus p
33	98	2.3	249	26	W23008	Canine herpesvirus un
34	96	2.3	420	13	R75754	BAV3 E1B 56K protein.
35	97	2.3	421	26	W34556	Staphylothermus marin
36	97	2.3	421	36	W49860	Staphylothermus marin
37	96	2.3	584	3	R14670	Truncated poly Ig-rec
38	96	2.3	773	20	W03177	Rabbit poly-immunoglo
39	96	2.3	869	10	R53732	S. cerevisiae Plci pr
40	96	2.3	883	14	R71415	E. coli PEPC (wild-ty
41	98	2.3	936	9	R52579	Recombinant collagena
42	98	2.3	1008	33	W63721	C. histolyticum ChCI
43	97	2.3	2721	2	P70647	Sequence of N-termina
44	94	2.2	411	21	W12379	P300-CelB fusion cons
45	94	2.2	412	21	W12380	P300-CelB fusion cons

ALIGNMENTS

RESULT 1
ID W00383 standard; Protein; 574 AA.
AC W00383;
DE 31-JAN-1997 (first entry)
DE Bacillus cellulase BCE 113.
KW Cellulase; BCE 113; detergent; surfactant; laundry;
KW tensile strength; antipilling
OS Bacillus sp. strain CBS 669.93.
FH Key Location/Qualifiers
FT peptide 1..26
FT protein /label= Sig_peptide
FT protein /label= Mat_protein
FT /note= "the 63 kDa mature protein is used in
detergent compns."

W09634092-A2
31-JAN-1997

26-APR-1996; E01755.
28-APR-1995; EP-201115.
PR 12-MAR-1996; US-614115.
PA (GENM) GENENCOR INT INC.
PI Kottwitz B, Lenting HBM, Maurer K, Van Beckhoven RFWC;
PI Van Solingen P, Weiss A;
DR WPI: 96-497624/49.
DR N-PSDB; T41849.

Cellulase with low ratio of tensile strength loss to antipilling
properties - used in detergent composition which provides
anti-greying, softening, anti-wrinkling and colour protection to
fabrics
Claim 10; Fig 4; 33pp; English.
CC A novel cellulase (W00383), designated BCE 113, of alkali-tolerant
CC Bacillus sp. strain CBS 669.93 has a tensile strength loss to
CC antipilling rate below 1% it can be isolated from CBS 669.93
CC fermentation broth or expressed at high levels in transformed host
CC cells utilising an isolated gene sequence (T41849). Cellulase BCE
CC 113, and similarly isolated cellulase BCE 103 (see also W00382),
CC show good activity at alkaline pH, and can be used in laundry
CC detergent compns. to provide anti-greying, softening, anti-
CC wrinkling and colour protection to fabrics.
SQ Sequence 574 AA;

Query Match 100.0%; Score 4235; DB 19; Length 574;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00; Mismatches 0; Indels 0; Gaps 0;
 Matches 574; Conservative 0

Db 1 mkwmksmvlavlvsvfapavssanedvklldiqsvvrdmqgwnlgnftdavgdgt 60
 |||||
 QY 1 MKWKSVMVLAVLVSVFAPAVSSANEDVKTLDIQSVVRDMPQGNLGNFTDVGQDET 60
 |||||

Db 61 awgnprvtrelrieriadeqysiripvtwengrsgadypidpqlnrdevvqwaaleed 120
 |||||
 QY 61 AWGNPRVTRELIERIADEQYSIRIPVTWENRIGGADYPIDPQFLNRDEVVQWALEED 120
 |||||

Db 121 lyminlhhsdswliwemehnyngmakysrlweqlshhfkdyptkImfesynepkfsqn 180
 |||||
 QY 121 LYMINLHHSWSLWIWEMEHNYNGMAKYSLWEQLSHHFKDYPTKLMFESVNEPKFSQN 180
 |||||

Db 181 wgeirenhalldlntvffveivrgsgqndirplvptmetatsqpllnnlyqtidkld 240
 |||||
 QY 181 WGEIRENHALLDLNLTVFFVEIVRSGSQNDIRPLVPTMETATSQPLLNLYQTIDKLD 240
 |||||

Db 241 dplnatvhygfwfsvniagytfrfeedskreietfdrvhhtfvargipvlgfql 300
 |||||
 QY 241 DPNLIATVHYGFWFVSNIAGYTFRFEEDSKREIETFDVRVHTFVARGIPVLGFQL 300
 |||||

Db 301 gfdkhtgvioggeklkffeylihlhlnrdithmlwdngqghfnrhtyewydeelfdmrlas 360
 |||||
 QY 301 GFDKHTGVIQGEKLFKFFEYLIHLHNRDITHMLWDNGQGHFNRTYEWYDEELFDMRLAS 360
 |||||

Db 361 wgrssvaesfnylkggdriadatvliqhgnetlglqangqrtpqgdyelngerltv 420
 |||||
 QY 361 WGRSSVAESFNYLKGGDRIADATVLIQHGNETLGLQANGQRTPQGQDYELNGERLTV 420
 |||||

Db 421 kahvlsaiagsgtlgtngmvtcaefnrgadwhfrvntvrtplvqstqghvsnfsipasnfg 480
 |||||
 QY 421 KAHVLSAIAAGSGTLGTNGMVTAEFNRGADWHFRVNTVTPVLSQTQGHVSNFSIPASNFG 480
 |||||

Db 481 nslatmeavydngnagpddwtskfeqyafspysydttheiklteaffrevrdgevrftfh 540
 |||||
 QY 481 NSLATMEAVYDNGNAGPDDWTSKFEQYAFSPSYDTHETIKLTAFFREVDRGEVRFTFH 540
 |||||

Db 541 fwsgeivnytiikngnqvgtgiaaqttsknknk 574
 |||||
 QY 541 FWSGEIVNYTIKNGNQVTGIAAQTTSKNKNK 574
 |||||

RESULT 2
 ID R13228 standard; Protein; 566 AA.
 AC R13228;
 DT 14-OCT-1991 (first entry)
 DE Endoglucanase encoded by endo2 gene.
 OS Cellulase activity; detergent.
 FH Bacillus spp. NCIMB 40250.
 FT Key Location/Qualifiers
 FT cleavage_site 30..31
 FT peptide 1..30
 FT /label= signal peptide
 FT /label= mature peptide
 FT W09110732-A.
 PN 25-JUL-1991.
 PD 18-JAN-1991; DK0013.
 PR 19-JAN-1990; DR-000164.
 PA (NOVO) NOVO NORDISK A/S.
 PI Jorgensen PL, Schulein M, Hansen C;
 DR N-PSDB; Q13002.
 DE Enzyme exhibiting cellulase activity from Bacillus sp. - is an
 PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
 PT fabrics.
 PS Claim 1; Page 80; 96pp; English.
 CC The enzyme is produced by transforming plasmid pPL382 to B.subtilis
 CC DnJ85 spp. and exhibits an endoglucanase activity of at least 10

CC (pref. at least 25) carboxymethyl cellulose (CMC) endoase units per
 CC mg total protein under alkaline conditions. It is especially useful
 CC as a cellulosytic agent and has been found to be more stable during
 CC washing (60 mins. at 40 deg.) in the presence of conventional
 CC detergents than a commercial cellulase preparation. It may also
 CC show increased storage stability in liq. detergents contg.
 CC proteases. The sequence was deduced from the DNA (Q13002), it is a
 CC product of the endo2 gene. See also R13227 and R13229.
 SQ Sequence 566 AA;

Query Match 60.4%; Score 2557; DB 3; Length 566;
 Best Local Similarity 57.6%; Pred. No. 9.20e-208;
 Matches 326; Conservative 117; Mismatches 119; Indels 4; Gaps 4;

Db 2 kkrsskvlslavllavlaavepnaalaaappsamqsyveamqpgwnlgnslavgade 61
 |||||
 QY 1 MKWKSVMVLAVLVSVFAPAVSSAN-EDVKTLDIQSVVRDMPQGNLGNFTDVGQDE 59
 |||||

Db 62 tlargnpritkellqnaaqysiripvtwdshigaapnygieaaylnrvgevvqwal 121
 |||||
 QY 60 T-ANGNPRVTRELIERIADEQYSIRIPVTWENRIGGADYPIDPQFLNRDEVVQWALE 118
 |||||

Db 122 anlyminhhsdswliwiskmesqhdqlarynaiwtgiantkfnspsklmfesyneprft 181
 |||||
 QY 119 EDLIVMINLHHSWSLWIWEMEHNYNGMAKYSLWEQLSHHFKDYPTKLMFESVNEPKFS 178
 |||||

Db 182 dg-gttdaekqkmldeinvssfnivrnsgqgnatrpvlstleasptqermtalynmt 240
 |||||
 QY 179 QNWGEIRE-NHHALLDLNLTVFFVEIVRSGSQNDIRPLVPTMETATSQPLLNLYQTID 237
 |||||

Db 241 klnknliahtvhygfwfsvniagytfrfeedskreietfdrvhhtfvargipvlgf 300
 |||||
 QY 238 KLDOPNLATVHYGFWFVSNIAGYTFRFEEDSKREIETFDVRVHTFVARGIPVLGF 297
 |||||

Db 301 gllfdkntgvioggeklkffeylihlhlnrdithmlwdngqghfnrhtyewydeelfdm 360
 |||||
 QY 298 GLLGFDKHTGVIQGEKLFKFFEYLIHLHNRDITHMLWDNGQGHFNRTYEWYDEELFDM 357
 |||||

Db 361 kasvtgrsstassdlihwkgtavkdtstvqlnlnqntitsvngttiksgtdytlnsr 420
 |||||
 QY 358 RASWGRSSVAESFNYLKGGDRIADATVLIQHGNETLGLQANGQRTPQGQDYELNGER 417
 |||||

Db 421 ltfkasqltkitslgkignvativtkfnrgadwhfrvntvrtplvqstqghvsnfsipasnfg 480
 |||||
 QY 418 LTVKAHVLSAIAAGSGTLGTNGMVTAEFNRGADWHFRVNTVTPVLSQTQGHVSNFSIPAS 477
 |||||

Db 481 lngdqlatmeavydngnagpddwtskfeqyafspysydttheiklteaffrevrdgevrft 540
 |||||
 QY 478 FNGNSLATMEAVYDNGNAGPDDWTSKFEQYAFSPSYDTHETIKLTAFFREVDRGEVRL 537
 |||||

Db 541 kqfswgeivnytiikngnqvgtgiaaqttsknknk 566
 |||||
 QY 538 TFWWSGEIVNYTIKNGNQVTGIAA 563
 |||||

RESULT 3
 ID R08199 standard; protein; 406 AA.
 AC R08199;
 DT 01-MAR-1991 (first entry)
 DE Neutral cellulase gene product.
 KW Cellulose.
 OS Ruminococcus albus.
 FH Key Location/Qualifiers
 FT misc_difference 44..58
 FT /label= N-terminal deletion
 FT protein 44..406
 FT J02265486-A.
 PN 30-OCT-1990.
 PD 07-APR-1989; 086714.
 PR 07-APR-1989; JP-086714.
 PA (SHIMIZU) SHIMIZU S.
 PI Shimizu S;
 DR WPI; 90-366319/49.

Db 108 ngafvinllhhethnhafs-e-tldtakeillekiwsgiaeeafxkdydehlifeglnepnrkn 165
 : ::::|::|::| : | : : : |::| : |::| : |::| : |::| : |::| : |::| :
 Qy 119 EDLYVMINLHSDSLWIYEMEHNNVGVMAKYSRLWEQLSNHFNDYPTKLMEFSVNKP-KF 177

 Db 166 dtpvewtgqdegwdav-nannnavflktirssggnnpkrhlmlppyaacnensfkn-f- 222
 : : : : |::| : |::| : |::| : |::| : |::| : |::| : |::| : |::| :
 Qy 178 -SQ-NW-GEIRENHHALLDLNTVFVEIVRQSGGQNDRPLVLPTMTETATSOPLLNLNQ 234

 Db 223 ifpe-dddkviasvhaypnfalnnngagadvkfdaagkkdlewlnlmkmkrfvdgqipm 281
 : : : : |::| : |::| : |::| : |::| : |::| : |::| : |::| : |::| :
 Qy 235 TIDKLDPNLIATVHYVGFWPFSVNI-AGYT-RFEEDSKREIIEITFDVRVHTTVAKGIPV 292

 Db 282 ilgygam--nrdd--ee-eratwaefymekvtamgvqvwwdng 321
 : ||::| : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 293 VLGEGFLGDPKGTGVTIQOGEKLFEEYLTHLNERNDIRITHMLWDNG 338

 RESULT 5
 ID R47496 standard; Protein; 800 AA.
 DT R47496;
 AT 21-JUL-1994 (first entry)
 DE Translated sequence of domains I and II of celd cDNA in clone pcNP4.
 KW Cellulase; celd; pcNP4; anaerobic rumen.
 OS Neocallimastix patriciarum.
 FH Key Location/Qualifiers
 FT region 1..37
 FT /label= N-terminus of beta-galactosidase alpha-peptide
 FT misc_difference 38..42
 FT /note= "derived from 5' oligo linker"

 W09A00578-A.
 PF 06-JAN-1993; AU0307
 PE 24-JUN-1993; AU0307
 PR 24-JUN-1992; AG-003096.
 PA (GSIR) COMMONWEALTH SCI & IND RES ORG.
 PI Aylward JH, Gobius KS, Orpin CG, Xue GP;
 DR WPI; 94-Q26214/03.
 NR NP-SDB; Q55036.
 PT Cloning of cellulase clones from anaerobic rumen - by isolating
 PT mRNA from culture, converting to cDNA etc. fungi, producing
 PT enzymes useful in food processing etc., and DNA for modifying
 PT rumen or silage bacteria
 PS Claim 10; Page 41-45; 71pp; English.
 CC Clone pcNP4.1 encodes celd. It was derived as follows. N.
 CC patriciarum was used to construct a cDNA library in ZAPII (in E.
 CC coli). Transformants were selected for enzyme activity using
 CC selective media. 11 colonies were positive, and of these 10 had the
 CC same restriction pattern, and the longest of these was designated
 CC celd (pcNP4.1)(Q55036)(R47496,R49102). A similar process was used
 CC to isolate the xylanase clone pNX-Tac (Q55037,R47497). An enzyme
 CC composition contg. celd and xylanase is claimed.
 SQ Sequence 800 AA;

Query Match 12.5%; Score 531; DB 9; Length 800;
 Best Local Similarity 32.9%; Pred. No. 1.49e-32;
 Matches 83; Conservative 72; Mismatches 85; Indels 12; Gaps

Db 84 etcwnpkttedmfklmdnqfnvfripttwsghfgapdykinekwlkrvhaivdyypk 143
 ||||| : : : : |::| : |::| : |::| : |::| : |::| : |::| :
 Qy 59 ETANGPRVTRIELRIADSGYKSIRTPVTWENRIGAPPDYPDPQLNRVDENVGWALE 118

 Db 144 ngafvinllhhethnhafs-e-tldtakeillekiwsgiaeeafxkdydehlifeglnepnrkn 201
 : : : : |::| : |::| : |::| : |::| : |::| : |::| : |::| : |::| :
 Qy 119 EDLYVMINLHSDSLWIYEMEHNNVGVMAKYSRLWEQLSNHFNDYPTKLMEFSVNKP-KF 177

 Db 202 dtpvewtgqdegwdav-nannnavflktirssggnnpkrhlmlppyaacnensfkn-f- 258
 : : : : |::| : |::| : |::| : |::| : |::| : |::| : |::| : |::| :
 Qy 178 -SQ-NW-GEIRENHHALLDLNTVFVEIVRQSGGQNDRPLVLPTMTETATSOPLLNLNQ 234

 Db 259 ifpe-dddkviasvhaypnfalnnngagadvkfdaagkkdlewlnlmkmkrfvdgqipm 317
 : : : : |::| : |::| : |::| : |::| : |::| : |::| : |::| : |::| :
 Qy 235 TIDKLDPNLIATVHYVGFWPFSVNI-A-GYT-RFEEDSKREIIEITFDVRVHTTVAKGIPV 292

Db 318 ilgeygamr:dn 329
:||||:|:|:
QY 293 VLGEFGLGFDK 304

RESULT

ID W56742 standard; Protein; 471 AA.

AC W56742; 1998 (first entry)

DE Orpinomyces cellulase celB.

KW Cellulase; endoglucanase; cellobiohydrolase; celB.

OS Orpinomyces sp. strain PC-2.

PN W09814597-A1.

PD 09-APR-1998.

PF 03-OCT-1997; U18008.

PR 04-OCT-1996; US-027883.

PA (UYGE-) UNIV GEORGIA RES FOUND INC.

PI Chen H, Li X, Ljungdahl LG;

DR WPI: 98-240096/21.

DR N-PSDB; V29477.

PT New recombinant DNA encoding Orpinomyces cellulase protein - useful for, e.g. producing recombinant Orpinomyces cellulase in host cell

PS Claim 1; Page 38-40; 69pp; English.

CC This polypeptide comprises cellulase celB of Orpinomyces sp. strain

CC CC-2, an anaerobic bovine rumen fungus Orpinomyces sp. PC-2. Its

CC amino acid sequence was deduced from an isolated cDNA clone (see

CC V29477). CelB has cellobiohydrolase activity, with highest

CC activity at pH 5.2-6.2 and 50 degC. CelA (see W56738) and celC

CC (see W56739) cellulolytic enzymes of Orpinomyces sp. PC-2 are also

CC provided. Recombinant DNA molecules encoding Orpinomyces cellulase

CC proteins are claimed, as well as recombinant cells selected from

CC Saccharomyces cerevisiae, Escherichia coli, Aspergillus,

CC Trichoderma reesei, Pichia, Penicillium, Streptomyces or Bacillus,

CC and a method for producing recombinant cellulase by culturing these

CC host cells.

CC Sequence 471 AA;

SQ

Query Match

Best Local Similarity 12.2%; Score 517; DB 32; Length 471;

Matches 104; Conservative 99; Mismatches 123; Indels 24; Gaps 19;

Db 1 mkfinslslglviagcaamrnis-kelvkeltigwslgntldascvetnyskdtas 59

QY 1 MKWKSMVLAWLVLVGFVAPAVSANEDVKTLDIQ-SYVRDMQPGWNLGNTFD-AVGQD 58

Db 60 etcgvnvtktgelykisdqfntfrpttswghfgdpydkisdvwmkrvhevvdvaln 119

QY 59 ETANGNPRVTRILLERIADEGYSIRIPVTWENRIGGAPDPIDPQFLNRVDEVVQWALE 118

Db 120 tggayalinhethw--nyafqknlesakkilvakiwqtaefgydehlifegmneprkv 177

QY 119 EDLYVMINLHSDSLWIYEMEHNYNGVMKYSLWEQLSNHFKDYPYTKLMFESVNEP-KF 177

Db 178 gdpawtgqdegvn-fvnemalfvktiratggnanrhlmityaasvndgsinn-fk 235

QY 178 SQ--NW-GEIRENHALLDLNTVFEIVROSGQNDIRPLVLTMETATSQPLLNLYQ 234

Db 236 ypng-dd-kvilslyhsypynfalnpggalsnfy-dgn-eidwvmtinsfskigpiv 291

QY 235 TIDKLDPNLITVHYVGFVWPFVNIA-G-YTRFEEDSKREIITFDVHHTFVARGIPV 292

Db 292 lgef--vamrdrn---ed-drerwqeyikkatalgipcvlwdng-yfe 334

QY 293 VLGEFGLGFDKHGVYIQGEKLAFFELIHLNDRDITHMLWNGQHFN 342

RESULT

ID W34566 standard; Protein; 360 AA.

AC W34566; 1998 (first entry)

DE Thermotoga OCl/4V endoglucanase.

KW Glycosidase; thermostable; textile; food processing; pharmaceutical;

QY

KW detergent; baking; industry; Thermococcus; Staphylothermus;
OS Pyrococcus; glucose; soluble oligosaccharide; endoglucanase.
PN Thermotoga sp.
PD W09723417-A1.
PF 17-JUL-1997.
PR 10-JAN-1997; U00092.
PR 13-SEP-1996; US-712612.
PR 11-JAN-1996; US-583787.
PA (RECO-) RECOMBINANT BIOTRANSFORMATION INC.
PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
DR WPI: 97-372858/34.
DR N-PSDB; T93690.
PT New thermostable glycosidase(s) - from Thermococcus, Staphylothermus
PT and Pyrococcus, used in the textile, food processing,
PT pharmaceutical, detergent and baking industries
PS Claim 4; Fig 13; 82pp; English.
CC The present sequence represents endoglucanase isolated from Thermotoga.
CC The enzyme or its encoding nucleic acid sequence is used for generating
CC glucose from soluble oligosaccharides. The enzyme can be used in the
CC food processing, pharmaceutical, textile, detergent and baking
CC industries. The enzyme is also used to treat lactose intolerance, as a
CC diagnostic reporter molecule, in corn wet milling or in the fruit juice
CC industry. The enzymes can be used to hydrolyse guar gum to remove
CC non-reducing terminal mannose residues. The nucleic acids encoding the
CC enzyme may be used to generate probes to identify similar sequences.
SQ Sequence 360 AA;

Query Match 7.9%; Score 335; DB 26; Length 360;

Best Local Similarity 42.3%; Pred. No. 1.82e-16;

Matches 55; Conservative 28; Mismatches 37; Indels 10; Gaps 6;

Db 62 nignales-pfeagwv-rieheylfeinkrgfdsvripirwsahisekpydyidrnfl 118

QY 47 NLGTFDVGQDETANGPNRVTRELIERIADGYSKIRIPVTWENRIGGAPDPIDPQFL 106

Db 119 ervnhvdralennltviinthfee-l--ygepdkygdvive---ivrqakfkdype 172

QY 107 NRVDVQWALEEDLYVMINLHH-DSWLWYEMEHNYNGVMKYSLWEQLSNHFKDIPY 155

Db 173 nlffeinyep 182

QY 166 KLMFESVNEP 175

RESULT

ID W49870 standard; Protein; 360 AA.

AC W49870;

DT 21-DEC-1998 (first entry)

DE Thermotoga OCl/4V endoglucanase.

KW Glycosidase; OCl/4V; thermostable enzyme; oligosaccharide; glucose;

QY sugar; baking; textile; detergent; endoglucanase.

OS Thermotoga sp. strain OCl/4V-33GPI.

PN W09824799-A1.

PD 11-JUN-1998.

PF 08-DEC-1997; U22623.

PR 10-OCT-1997; US-949026.

PR 06-DEC-1996; US-056916.

PA (DIVE-) DIVERSA CORP.

PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;

DR WPI: 98-362407/31.

DR N-PSDB; V36919.

PT Glycosidase enzymes from organisms of the genera Staphylothermus,

PT Pyrococcus and Thermococcus - for deriving sugar from

PT oligosaccharides, useful in the e.g. food processing, textile or

PT baking industries

PS Claim 1; Fig 13a-b; 92pp; English.

CC This is the amino acid sequence of endoglucanase OCl/4V, deduced

CC from a polynucleotide (see V36919) of a Thermotoga sp. OCl/4V clone

CC (33GPI) that grows optimally at 75 degC. The sequence shows 65%

CC amino acid identity to an endo-1,4-beta-endoglucanase of

CC Caldocellum thermoceillum. The invention provides 18 polynucleotides

CC (see V36907-24) coding for thermostable glycosidases (see W49858-75)

CC having glucosidase, alpha-galactosidase, beta-galactosidase,

KW	Endoglycoceramidase; glycolipid; sugar chain engineering.
OS	Rhodococcus sp. M-777.
FH	Key Location/Qualifiers
FT	peptide 1..29
FT	/label= Signal_peptide
FT	protein 30..490
FT	/label= Mature_protein
EP	-751222-A2.
PN	02-JAN-1997.
PD	28-JUN-1996; 110523.
PF	29-JUN-1995; JP-188465.
PR	(TAKI) TAKARA SHUZO CO LTD.
PPA	Ito M, Izu H, Izumi Y, Kato I, Kurome Y, Sano M;
PI	WPI; 97-054679/06.
DR	N-PDSB; T50943.
DR	New DNA encoding endo-glyco-ceramidase - useful in structural and
PT	functional analysis of glyco-lipid(s)
PFS	Claim 1; Page 11-13; 25pp; English.
PSP	The sequences given in W0209-10 represent full length and mature
CC	endoglycoceramidase respectively, which is useful in structural
CC	and functional analysis of glycolipids in sugar chain engineering.
CC	Recombinant endoglycoceramidase can be produced with high purity at
CC	relatively low cost using the DNA sequences encoding these proteins.
CC	Sequence 490 AA;
SQ	
Query Match 3.7%; Score 156; DB 23; Length 490;	
Best Local Similarity 29.8%; Pred. No. 1.04e-02;	
Matches 34; Conservative 31; Mismatches 40; Indels 9; Gaps 9	
Db	29 aaagsgsgsgtal-tpsyldk-ddgrslrlrgfntassaksapdgmppqfteadlareyd 86 : ::: : ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
QY	21 PAVSSANEDVKTLIDIQSYVRDMQPGNLNG-NTFDAVGQDETAW-GNPRVTR-ELIERIAD 77 : ::: : ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db	87 mgtnfvrflwrs-vepagg-vydcqylidrvedrvgyaaerg-ykvmdmhgd 137 : ::: : ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
QY	78 EGYKSIRIPWTENRIGGAPDYPDFQLNRVDVVQVALEEDLY-VMINLHHDD 130 : ::: : ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
RESULT 11	
ID W57421 standard; Protein; 418 AA.	
AC W57421;	
DE 02-SEP-1998 (first entry)	
KW Amino acid sequence of the specification.	
KE Regulatory sequence; cellulase cbhl gene; mass production;	
KW Humicola insolens; endo-glucanase NCE4.	
OS Trichoderma viride.	
FH Key Location/Qualifiers	
FT Peptide 1..21	
FT Protein 22..418	
FT /note= "signal peptide"	
FT /note= "mature protein"	
FN W09811239-Al.	
PN 19-MAR-1998.	
PD 16-SEP-1997; J03268.	
PR 13-SEP-1996; JP-243695.	
PA (MEIJ) MEIJI SEIKA KAISHA LTD.	
PI Aoyagi K, Moriya T, Murakami T, Sumida N, Watanabe M;	
DR WPI; 98-250959/22.	
DR N-PDSB; V29597.	
PT Regulatory sequence for Trichoderma viride derived cellulase cbhl	
PT gene - for producing Humicola insolens derived endo-glucanase	
PS Disclosure; Pages 48-51; 92pp; Japanese.	
CC The present sequence appears in the specification. The specification	
CC describes a new regulatory sequence for trichoderma viride derived	
CC cellulase cbhl gene and the establishment of a system for mass producing	
CC cellulase in moulds such as T. viride. As the regulatory sequence of	
CC cbhl genes originating in T. viride can highly express objective	
CC proteins, proteins such as cellulase can be expressed. An expression	
CC vector containing the regulatory sequence and Humicola insolens derived	
CC endo-glucanase NCE4 DNA was produced, and used to produce endo-glucanase	
CC at 15 grams per litre.	
CC Sequence 418 AA;	
SQ	

DE	T. longibrachiatum endoglucanase EGII.
KW	Cellulase; cellulose; signal; catalytic core; cellulase binding;
LK	linker.
OS	Trichoderma longibrachiatum.
FH	Key Location/Qualifiers
FT	peptide 1..21 /label= signal
FT	/note= "seq id no 32"
FT	domain 22..57 /label= cellulose binding domain
FT	/note= "seq id no 7"
FT	region 58..91 /label= linker
FT	/note= "seq id no 24"
FT	domain 92..418 /label= catalytic core domain
FT	/note= "seq id no 16"
PX	WO9516702-A.
PD	22-JUN-1995.
PF	19-DEC-1994; U14163.
PP	17-DEC-1993; US-169948.
PR	(GEMV) GENENCOR INT INC.
PA	Clarkson KA, Collier KD, Fowler T, Larenas E, Ward M; WPI; 95-231574/30.
DR	N-PDSB; Q91284.
DQ	Pure, truncated fungal cellulase protein from Trichoderma - useful to reduce or eliminate dye, colourant or pigment back-staining or redeposition in stone-washing or bio-polishing
PT	Claims 17, 41, 63, 64, 65; Figure 4; 105pp; English.
CC	Figure 4 depicts the genomic DNA and AA sequence of EGII derived from T. longibrachiatum. A truncated fungal cellulase which consists essentially of the AA sequence in SEQ ID NO. 8 is claimed.
CC	Also claimed are DNA gene fragments encoding SEQ ID NOS: 12, 20 and 16; 16: 16 and 20; and 16, 20 and 12.
CC	Genes for EG I and EG II have been isolated from T. longibrachiatum and the protein domain structure has been confirmed (Penttilä, M. et al., 1986, Gene, 45, 253-263; Van Arsdeell, J.N. et al., 1987, Bio/Technology, 5, 60-64; and Saloheimo, M. et al., 1988, Gene 63, 11-21).
CC	Sequence 418 AA;
SQ	
Query Match 3.3%; Score 141; DB 14; Length 418;	
Best Local Similarity 22.8%; Pred.No. 1.18e+01;	
Matches 23; Conservative 30; Mismatches 46; Indels 2; Gaps	
Db	142 vnedgmtifrlprgwylvnnlgnldstisksyqdlvggclsgaycivdihayar~w 200 : :: :: : : : : : : : : : :
QY	75 IADEGYKSIRIPTWNRRGGAPDYPIDFQLNRVDVVQWALEDLYMNLHHDSLWLW 134 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : ~
Db	201 nggilggpggttaftslsqaskyas~gsrvwfginneep 240 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : ~
QY	135 IYMEHNNGVMAYKSLWEQLSNHFKDPTKLMFESVNEP 175 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : ~
RESULT 14	
ID	W02032 standard; Protein; 418 AA.
AC	W02032:
DT	28-OCT-1996 (first entry)
DE	Trichoderma endoglucanase II.
KW	Endoglucanase II; EGII; cellulase; cellulose; denim;
KX	stonewashing; dye redeposition; backstaining.
OS	Trichoderma longibrachiatum.
FH	Key Location/Qualifiers
FT	peptide 1..21 /label= Sig_peptide
FT	/note= "seq id no 32"
FT	protein 22..418 /label= Mat_protein
FT	/note= "seq id no 7"
FT	domain 22..57 /label= Cellulose_binding_domain
FT	/note= "seq id no 24"
FT	region 58..91 /label= Linker_region
FT	/note= "seq id no 16"
FT	domain 92..418 /label= Catalytic_core_domain
FT	/note= "seq id no 32"

FT /note= "catalytic core domain is the preferred
FT domain for use in constructs of the
FT invention"

FN W09623928-A1.
PD 08-AUG-1996.
PF 29-JAN-1996; U00977.
PR 01-FEB-1995; US-382452.
PA (GEMV) GENENCOR INT INC.
PI Clarkson KA, Collier KB, Fowler T, Larenas E, Ward M;
DR WPI: 96-371466/37.
DR N-PSDB: T32223.
PT Treatment of cellulose-contg. fabrics such as denim, e.g.
PT stone-washing - using truncated cellulase enzyme to increase
PT abrasion and give reduced redeposition of dye
PS Disclosure: Fig 4A-4C; 124pp; English.
CC The amino acid sequences for Trichoderma longibrachiatum
CC cellobiohydrolase I (CBHI) (W02022), CBHII (W02025), endoglucanase I
CC (EGI) (W02029), EGII (W02032), and EGIII (W02034) were deduced from
CC the respective genomic DNA sequences (T32220-24). The CBHI, CBHII,
CC EGI and EGII enzymes have catalytic core domains useful for reducing
CC dye redeposition (backstaining) on cellulose-contg. fabrics such as
CC denim, whilst maintaining or increasing abrasion during stonewashing.
CC Truncated enzymes comprising these catalytic core domains can be obtd.
CC by proteolysis of the complete enzyme or by inserting the appropriate
CC DNA fragment into a vector, using this to transform a Trichoderma
CC sp. host cell, and recovering the recombinant core domain.
SQ Sequence 418 AA;

Query Match 3.3%; Score 141; DB 18; Length 418;
Best Local Similarity 22.8%; Pred. No. 1.18e-01;
Matches 23; Conservative 30; Mismatches 46; Indels 2; Gaps 2;
Db 142 vnedgmtifrlpvvgvlynnlgnldstskyskylvgcgclslgactvdihnyar-w 200
QY 75 IADGKYSIRIPVWENRIGGAPDPIDPQLNRVDEVVQWALEEDLYVMINLHDSWLW 134
Db 201 nggllgqggptnaqftslwslaskyas-qsrwvfgimnep 240
QY 135 IYEMEHNYNGVMKYSLSWQLSLNHFDPYTKLMFESVNEP 175

RESULT 15

ID R88407 standard; Protein: 429 AA.
AC R88407;
DT 05-JUL-1996 (first entry)
DE Beta-(1,6)-endoglucanase.
KW Trichoderma harzianum beta-(1,6)-endoglucanase; Aspergillus oryzae;
KW beta-glucan degradation; pustulanase.
OS Trichoderma harzianum.
PN W09531534-A1.
PD 23-NOV-1995.
PF 11-MAY-1995; DK0189.
PR 11-MAY-1994; DK-000547.
PA (NOVO) NOVO-NORDISK AS.
PI Andersen LN, Christgau S, Dalboge H, Kauppinen MS;
PI Kofod LV, Olsen HS;
DR WPI: 96-010921/01.
DR N-PSDB: T09881.
PT DNA encoding beta-1,6-endoglucanase from Trichoderma harzianum -
PT useful, e.g., in prepn. of yeast extracts, as antifungal agent, in
PT cleaning compns., etc.
PS Claim 1; Page 35; 49pp; English.
CC Trichoderma harzianum (CBS 243.71) beta-(1,6)-endoglucanase may be
CC produced recombinantly, and has a molecular weight of about 50,000,
CC an apparent isoelectric point of 5.6, an optimal temperature of
CC 30-40 deg, an optimal pH of 4-6, preferably 5, and a specific
CC activity of about 100 U/mg. The enzyme, or compositions enriched
CC in it, are used to modify or degrade beta-glucans, particularly for
CC rupturing or lysing cell walls of microorganisms thereby enabling
CC recovery of desirable products produced by the microorganisms.
CC Typical applications include the production of protoplasts and
CC yeast extracts, use in wine and press juice making, as fungicides,
CC to remove excess dye from textiles, to remove moulds on coatings or

CC biofilms from surfaces, for cleaning dentures and removing plaque,
CC and in the extraction of mannoproteins from microbial cell walls.
SQ Sequence 429 AA;
Query Match 2.8%; Score 119; DB 17; Length 429;
Best Local Similarity 32.8%; Pred. No. 3.64e+00;
Matches 22; Conservative 17; Mismatches 24; Indels 4; Gaps 4;
Db 111 winpatvqsvhdvlgltiripigysynaivdtasepfadgnlqlpyldavvqkaedlgi 170
QY 66 RVTRIELIERIADEGYSIRIPVT-WE-NRIGGAPDPI-DPOF-LNRVDEVVQWALEEDL 121
Db 171 yviidlh 177
QY 122 YVMINLH 128
Search completed: Fri Jun 4 10:07:48 1999
Job time : 121 secs.

W P S R E H
(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

psrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Jun 4 10:04:13 1999; Maspar time 24.78 Seconds
928.039 Million cell updates/sec
Tabular output not generated.

Title: >US-08-945-574-2
Description: (1-574) from US08945574.pep
Perfect Score: 4235
Sequence: 1 MKWMKSMWLVVSVFVA.....GNQVTGIAQTTSKNNKK 574

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60

1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 51.634; Variance 110.240; scale 0.468

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	2570	60.7	566	2 JH0218	cellulase (EC 3.2.1.4	0.00e+00
2	590	15.3	441	2 A44815	cellulase (EC 3.2.1.4	1.90e-100
3	566	15.7	515	2 S20493	endoglucanase - Clost	6.40e-96
4	660	15.6	814	1 C2CLEM	cellulase (EC 3.2.1.4	8.63e-95
5	614	14.5	364	2 S12017	endoglucanase A - Rum	3.70e-86
6	603	14.2	409	2 S12018	endoglucanase B - Rum	4.21e-84
7	597	14.1	406	2 A43722	cellulase (EC 3.2.1.4	5.55e-83
8	528	12.5	482	2 J03002	cellulase (EC 3.2.1.4	3.54e-70
9	525	12.4	473	2 S40507	endoglucanase - rumen	1.26e-69
10	510	12.0	517	2 I40798	cellulase (EC 3.2.1.4	7.29e-67
11	489	11.5	547	2 JQ0356	cellulase (EC 3.2.1.4	5.16e-63
12	479	11.3	475	1 C2CLCA	cellulase (EC 3.2.1.4	3.47e-61
13	470	11.1	455	2 S16559	cellulase (EC 3.2.1.4	1.52e-59
14	447	10.6	363	2 I40234	carboxymethylcellulas	2.27e-55
15	422	10.0	584	2 JQ1229	cellulase (EC 3.2.1.4	7.34e-51
16	412	9.7	900	2 JH0157	cellulase (EC 3.2.1.4	4.57e-49
17	324	7.7	580	2 S22458	cellulase (EC 3.2.1.4	1.44e-33
18	304	7.2	584	2 S27500	xylanase - Prevotella	3.88e-30
19	139	3.3	341	2 JCI1201	carboxymethylcellulas	2.41e-04
20	141	3.3	418	1 S28372	cellulase (EC 3.2.1.4	1.31e-04
21	141	3.3	419	2 S23225	glucan 1,3-beta-glucos	1.31e-04
22	137	3.2	410	1 S68153	cellulase (EC 3.2.1.4	4.43e-04
23	126	3.0	338	2 JCI115	endoglucanase precurs	1.15e-02

24	127	3.0	764	2 S50878	hypothetical protein	8.63e-03
25	119	2.8	159	2 S72326	glucan 1,3-beta-glucos	8.42e-02
26	119	2.8	1848	2 A44140	cellulose-binding pro	8.42e-02
27	115	2.7	430	2 S55325	endo-beta-1,6-glucana	2.54e-01
28	113	2.7	597	2 D70100	phosphoglucosyltransf	4.37e-01
29	112	2.6	264	2 S41113	dnaE protein - Entero	5.71e-01
30	109	2.6	516	2 F71529	probable exodexyribon	1.27e+00
31	109	2.6	570	2 S56132	cellulase (EC 3.2.1.4	1.27e+00
32	104	2.5	181	2 D70005	conserved hypothetical	4.60e+00
33	105	2.5	304	2 A49496	malate dehydrogenase	3.56e+00
34	105	2.5	362	2 C64807	ybgO protein - Escher	3.56e+00
35	106	2.5	481	2 B71929	glutamine synthetase	2.76e+00
36	106	2.5	481	2 H64583	glutamine synthetase	2.76e+00
37	104	2.5	542	2 JN0438	carboxylesterase (EC	4.60e+00
38	105	2.5	567	2 F64453	oxaloacetate decarbox	3.56e+00
39	107	2.5	607	2 S49528	endoxylanase - rumen	2.13e+00
40	107	2.5	607	2 S24754	endo-1,4-beta-xylanas	2.13e+00
41	106	2.5	608	2 T02299	hypothetical protein	2.76e+00
42	104	2.5	880	2 S44833	F54H12.5 protein - Ca	4.60e+00
43	107	2.5	1802	2 T00020	bacterial blight-resi	2.13e+00
44	103	2.4	400	2 B70092	serine proteinase Do	5.91e+00
45	103	2.4	986	2 S12021	thermoactive cellulase	5.91e+00

ALIGNMENTS

RESULT 1
ENTRY
TITLE

JH0218 #type complete
cellulase (EC 3.2.1.4) precursor - Bacillus lautus (strain
PL236)

ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM #formal name Bacillus lautus
DATE 12-Feb-1993 #sequence revision 12-Feb-1993 #text_change
12-Nov-1998

ACCESSIONS

REFERENCE JH0218

#authors Jorgensen, P.L.; Hansen, C.K.

#journal Gene (1990) 93:55-60

#title Multiple endo-beta-1,4-glucanase-encoding genes from Bacillus

#cross-references M101303016

#accession JH0218

#molecule_type DNA

##residues 1-566 ##label JOR

##cross-references GB:M33762; NID:g142870; PID:g142871

COMMENT The ability to hydrolyze crystalline cellulose of B. lautus PL236
is related to its production of at least four kinds of the
endo-beta-1,4-glucanases.

GENETICS

#gene

FUNCTION

#description

hydrolysis of 1,4-beta-D-glucosidic linkages in

beta-D-glucans such as cellulose and lichenan; can

hydrolyze such linkages in beta-D-glucans that also contain

1,3-linkages

#pathway

cellulose degradation

CLASSIFICATION

#superfamily cellulase CCA; Clostridium cellulase repeat

homology

KEYWORDS

extracellular protein; glycosidase; hydrolase; polysaccharide

degradation

FEATURE

1-30

#domain signal sequence #status predicted #label SIG\

#product cellulase #status predicted #label MAT\

#active_site Glu #status predicted

31-566

#length 566 #molecular-weight 62621 #checksum 3990

SUMMARY

Query Match 50.73; Score 2570; DB 2; Length 566;

Best Local Similarity 57.88; Pred. No. 0.00e+00;

Matches 327; Conservative 116; Mismatches 119; Indels 4; Gaps 4;

DB 2 KRRSRKVLISLAVVALLAAVFPNAAAPPASQSYVEAMQPQWNLGNSLDVAGDE 61

QY 1 MKWMKSMWLVVSVFVAPVSSAN-EDVKTLDIQSYVRDMPQWNLGNTFDVAGQDE 59

[illegible]

ENTRY	#type complete
CZC1EM	cellulase (EC 3.2.1.4) E precursor - Clostridium thermocellum

[illegible]

```

Db 115 MLPEAGFNLRIPIVSWGNHLIDN-NYITDPAWMDRVOEIVNYGIDDDGMYVILNTHHEW- 172
QY 74 RIADEGYKSIPIVPTWENRIGGAPDYPIDPQFLNRVDEVVQWALEEDLYVMINLHDSWL 133
Db 173 YM-PKPEKODGIEELKAIWSQIADRFKGYDEHLIFEGLEPRLRGEAGWTGTSEARE- 230
QY 134 WIYEMEHNYNGVAKYRSLEQSLNHFKDYPTKLMFESVNEPKF-SQ--NWGEIRENHHA 190
Db 231 IINEYKAFVETVRASGNGDRCLMI-T-GYAASSGY-NNL-SAIPELPEDSKLIISVH 286
QY 191 LLDLNTVFFVIRVQSGQNDIRPLVPTMETATSQPLNLLNYQTIDKLDNPN-LIAIVH 249
Db 287 AYLPSPALDTKGDYKDPEDT-AIPLFESLNELFISRDIPVIVGEGSNKKNIDRVR 345
QY 250 YYGWFPFSVNIAGYTRFEEDSKREIETFDVRVHTFVARGIPVVLGEGFLGDKHTGVI 309
Db 346 KCLDDYLGNAAKYDIPCVW-WDNVARI-GCNEFGLNRQEDYKFFPKLMDVFK 397
QY 310 Q-OGEKL-KFFEYLIHNLNERDTHMLWDNGQHF---NRHTYEWYDELFDMRLR 358

RESULT 7
ENTRY #type complete
TITLE cellulase (EC 3.2.1.4) precursor - Ruminococcus albus
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM #formal_name Ruminococcus albus
DATE 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change
13-Sep-1998
ACCESSIONS A43722; A39134
REFERENCE A43722
#authors Ohmiya, K.; Kajino, T.; Kato, A.; Shimizu, S.
#journal J. Bacteriol. (1989) 171:6771-6775
#title Structure of a Ruminococcus albus endo-1,4-beta-glucanase
#accession A43722
#cross-references MUID:90078126
#molecule_type DNA
#status preliminary
#residues 1-406 #label OHM
#cross-references GB:M30928; NID:g152638; PID:g152639
REFERENCE A39134
#authors Ohmiya, K.; Deguchi, H.; Shimizu, S.
#journal J. Bacteriol. (1991) 173:636-641
#title Modification of the properties of a Ruminococcus albus
#accession A39134
#cross-references MUID:91100351
#molecule_type DNA
#status preliminary
#residues 44-70 #label OH2
ACTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1,3-linkages
#pathway cellulose degradation
CLASSIFICATION #superfamily Ruminococcus albus cellulase
KEYWORDS glycosidase; hydrolase; polysaccharide degradation
SUMMARY #length 406 #molecular-weight 45389 #checksum 2035

Query Match 14.1%; Score 597; DB 2; Length 406;
Best Local Similarity 31.6%; Pred. No. 5.55e-83;
Matches 92; Conservative 85; Mismatches 96; Indels 18; Gaps 14;

Db 53 VPVSOTHTNDTMTVTSKDLVAKWNTGNLGMTDATAQGLSEVSWPLKVTNKMID 112
QY 19 VAPAVSSANDVKTLDIQSVRDMQPCWNLGNTFDVAGQD---ETANGNPRVRE--LIE 73
Db 113 MLPEAGFNLRIPIVSWGNHIDD-KYTSDFPAMWDRVOEIVNYGIDDDGMYVILNTHHEW- 170
QY 74 RIADEGYKSIPIVPTWENRIGGAPDYPIDPQFLNRVDEVVQWALEEDLYVMINLHDSWL 133

```

```

Db 171 YM-PKPEKODGIEELKAIWSQIADRFKGYDEHLIFEGLEPRLRGEAGWTGTSEARE- 228
QY 134 WIYEMEHNYNGVAKYRSLEQSLNHFKDYPTKLMFESVNEPKF-SQ--NWGEIRENHHA 190
Db 229 IINEYKAFVETVRASGNGDRCLMI-T-GYAASSAY-NNL-SAIPELPEDSKLIISVH 284
QY 191 LLDLNTVFFVIRVQSGQNDIRPLVPTMETATSQPLNLLNYQTIDKLDNPN-LIAIVH 249
Db 285 AYLPSPALDTKGDYKDPEDT-AIPELFELHNLFIKSGPIVIVGEGFTM 334
QY 250 YYGWFPFSVNIAGYTRFEEDSKREIETFDVRVHTFVARGIPVVLGEGFL 300

RESULT 8
ENTRY #type complete
TITLE cellulase (EC 3.2.1.4) - rumen fungus (Neocallimastix
frontalis)
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM #formal_name Neocallimastix frontalis
DATE 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change
05-Dec-1998
ACCESSIONS JE0302
REFERENCE JE0302
#authors Fujino, Y.; Ogata, K.; Nagamine, T.; Ushida, K.
#journal Biosci. Biotechnol. Biochem. (1998) 62:1795-1798
#title Cloning, sequencing, and expression of an endoglucanase gene
from the rumen anaerobic fungus Neocallimastix frontalis
MCH3.
#accession JE0302
#molecule_type mRNA
#residues 1-482 #label FUJ
#cross-references GB:U38843
GENETICS
#gene cels
KEYWORDS glycosidase; hydrolase
SUMMARY #length 482 #molecular-weight 54646 #checksum 7849

Query Match 12.5%; Score 528; DB 2; Length 482;
Best Local Similarity 30.8%; Pred. No. 3.54e-70;
Matches 88; Conservative 81; Mismatches 99; Indels 18; Gaps 15;

Db 78 ETCWGNPKTDEDMFKVLMNDQNFVRIPTTWSGHGFEAPDYKINEKWLKRVHIVDYK 137
QY 59 ETANGNPRVTRIELIERIADEGYKSIPIVPTWENRIGGAPDYPIDPQFLNRVDEVVQWALE 118
Db 138 NGAFVILNLHETWNHAFS-E-TLDTAKEILEKTSQIAKEFKDYDEHLIFGGLNEPKN 195
QY 119 EDLYVMINLHDSWLWIYEMEHNTYNGVNAKYRSLWEQLSNHFKYPTKLMFESVNEP-KF 177
Db 196 DTPVWETGGDQEGWDV-NAMNAVFLKTISSGGNPNKRHLMPYAAACNENSKN-F- 252
QY 178 -SQ-NW-GEIRENHALLDLDLNTVFFVIRVQSGQNDIRPLVPTMETATSQPLNLLNYQ 234
Db 253 IFPE-DDDKVTASVHAYAPYNFALNNGAGAVDKFADAGKDLKLEWINILMKKRFVDQIGPM 311
QY 235 TIDKLDNPNLTIATVHYIGFWPFSVNI-AGYT-RFEEDSKREIETFDVRVHTFVARGIPV 292
Db 312 ILGEYGM-NRDN---EE-PRATWAEFYMEKVTAMGPVQVWONG 351
QY 293 VLGEGFLGDFKHTGVIQQQGEKLFEEYLIHHLNERDITHMLWDNG 338

RESULT 9
ENTRY #type complete
TITLE endoglucanase - rumen fungus (Neocallimastix patriciarum)
ORGANISM #formal_name Neocallimastix patriciarum
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
24-Jul-1997
ACCESSIONS S40507
REFERENCE S40507
#authors Zhou, L.; Xue, G.; Orpin, C.G.; Black, G.W.; Gilbert, H.J.;
Hazlewood, G.P.
#journal Biochem. J. (1994) 297:359-364

```

#title	Intronless celB from the anaerobic fungus Neocallimastix patriciarum encodes a modular family A endoglucanase.									
#accession	S40507									
#status	preliminary									
#molecule_type	mRNA									
#residues	1-473 ##label ZHO									
SUMMARY	#length 473 #molecular-weight 53070 #checksum 1241									
Query Match	12.4%; Score 525; DB 2; Length 473;									
Best Local Similarity	29.6%; Pred. No. 1,26e-69;									
Matches	104; Conservative 103; Mismatches 118; Indels 26; Gaps 20;									
Db	1	MKFLNTFLLS	AAIIGSKAMKNISS--KELVKDITIGWLSGNTLDTACFETDLYNKQIA-	58						
Qy	1	MKKMKSMTVL	AAVLVVSFVAPAVSSANEDVKTLDIQ-SYVRDQMPG-WN-LGNTFDAQVQ	57						
Db	59	SETCHGNVKT	TOELYYKLSDLGNTFRIPRTTWSGHGPNADPYKINDQMKRVRHEIVDYAI	118						
Qy	58	DETAGNPRVT	RELIERIADEGYSKIRIPVTWENRIGGAPDYPIDPQFLNRVDEVVQWAL	117						
Db	119	NTGGVAILNI	HETWNAFOK--NLESAKKILVAIKWQATAAEFADYDEHILFEGMEPRK	176						
Qy	118	EDLVYMLNLH	SDSLWIEYEHNYNGWAKYRSLEQSLNHFKDYPTKLMFESVNEP-K	176						
Db	177	VGDPAEWNGD	YEGWN-FYNEKNDLFVKITRATGGNNALRHLMPYAAICINDGAINN-F	234						
Qy	177	FSQ--NW-GE	TRENHALLDDTLNTVFEIVRSGGQNDIRPLVPTMETATSOPLLNLY	233						
Db	235	K-FPSGDP-	KVIVSLHSYSFYNFALNNGAGISNFDGS--EIDWAMTINSKFISRGIP	290						
Qy	234	QTIDKLDPN	LATVHYIGFWFPSVNI-AG-YTRFEEDSKREIETFDVRVHHTFVARGIP	291						
Db	291	VIIGFEGAM-	-NRNN--ED-DRERWAEEYIKKATSIGVPCVTWDNG-YFE	334						
Qy	292	VYLGEFGLG	FDKHTGVIOQGEKLFYLIHHLNERDITHMLWDNGQHEN	342						
RESULT	10									
ENTRY	140798 #type complete									
TITLE	cellulase (EC 3.2.1.4) celA - Clostridium longisporum									
ALTERNATE_NAMES	endo-1,4-beta-D-glucanase									
ORGANISM	#formal_name Clostridium longisporum									
DATE	04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 21-Dec-1998									
ACCESSIONS	140798									
REFERENCE	140798									
#authors	Mittendorf, V.; Thomson, J.A.									
#journal	J. Gen. Microbiol. (1993) 139:3233-3242									
#title	Cloning of an endo-(1-->4)-beta-glucanase gene, celA, from the rumen bacterium Clostridium sp. ('C. longisporum') and characterization of its product, CelA, in Escherichia coli.									
#cross-references	MUID:94172316									
#accession	140798									
#status	preliminary; translated from GB/EMBL/DBJ									
#molecule_type	DNA									
#residues	1-517 ##label RES									
##cross-references	GB:L02868; NID:g144754; PID:g144755									
GENETICS										
#gene	celA									
CLASSIFICATION	#superfamily endoglucanase; bacterial cellulose-binding domain homology									
KEYWORDS	extracellular protein; glycosidase; hydrolase; polysaccharide degradation									
SUMMARY	#length 517 #molecular-weight 57660 #checksum 8670									
Query Match	12.0%; Score 510; DB 2; Length 517;									
Best Local Similarity	31.8%; Pred. No. 7,29e-67;									
Matches	91; Conservative 85; Mismatches 89; Indels 21; Gaps 17;									
Db	73	ETGWNPVTTK	AMIDIKKNAGFKTIRIPTTWGSHLDGN-N-KLNEEWKRVKEVYDCIA	130						
Qy	59	ETAWGNPRVT	RELIERIADEGYSKIRIPVTWENRIGGAPDYPIDPQFLNRVDEVVQWALE	118						

QY 284 TFVARGIPVVLGE 296

```
RESULT 12
ENTRY
TITLE C2CLCA #type complete
ALTERNATE_NAMES cellulase (EC 3.2.1.4) CCA precursor - Clostridium sp.
ORGANISM endo-1,4-beta-glucanase; endoglucanase-A
#formal_name Clostridium sp.
DATE 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change
ACCESSIONS JH0082
REFERENCE JH0082
#authors Faure, E.; Belaich, A.; Bagnara, C.; Gaudin, C.; Belaich, J.P.
#journal Gene (1989) 84:39-46
#title Sequence analysis of the Clostridium cellulolyticum endoglucanase-A-encoding gene, celCCA.
#cross-references MUID:90108715
#accession JH0082
#molecule_type DNA
#residues 1-475 #label FAU
#experimental_source ATCC 35319
#note a strong homology is found between the C-terminal duplicated region of this protein and regions in four endoglucanases and one xylanase from Clostridium thermocellum

GENETICS
#gene celCCA
#description catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-glucans such as cellulose and lichenin; can hydrolyze such linkages in beta-D-glucans that also contain 1,3-linkages
#pathway cellulose degradation
CLASSIFICATION #superfamily cellulase CCA; Clostridium cellulase repeat homology
KEYWORDS duplication; extracellular protein; glycosidase; hydrolase; polysaccharide degradation
FEATURE
1-26 #domain signal sequence #status predicted #label SIG\
26-475 #product cellulase #status predicted #label MAT\
415-438 #domain Clostridium cellulase repeat homology #label CCR\
446-469 #domain Clostridium cellulase repeat homology #label CCR
SUMMARY #length 475 #molecular-weight 53624 #checksum 5839

Query Match 11.38; Score 479; DB 1; Length 475;
Best Local Similarity 30.78; Pred. No. 3.47e-61;
Matches 107; Conservative 96; Mismatches 110; Indels 35; Gaps 26;

33 PNLQIPQKNIPNNNDGNFVKGRLGNLGNFTDFAFNCTNITNLDYETSWSGIKTKKMI 92
QY 21 PAVSSANEDVKTLDIOSYVRDMQPGWNLGNFTFA-----V-GQ-D-ETANGNPRVTRELI 72
DB 93 DAIRKQGFNTVRIPVSWHPHVSQS-DYKISDVMMNRVQEVVNYCIDNKMXYILNTHHDVD 151
QY 73 ERTADEGYSIRIPVTWENRIGGAPDPIDPQFLNRVDEVVQWALEEDLYVMINLHD-S 131
DB 152 KYGYPSPQYMASSKKYITSYVAQTAARFANYDEHLIFGCMNEPRLVGHANENWPELTN 211
QY 132 WLWIYEMEHNYNGVMARY-RSLWEQLSNHFKDYPYTKLMFESVNEPKF-SQ-N--WGIRE 186
DB 212 SDVVDVINCINOLNODFVNTRVATGKNASRYLMCPGY-VASPDGATNDYFRMPNDISGN 270
QY 187 NH--HAL--LDDLTNTVFFELVRSGGQNDIRPLVLPMTETATSOPLNNLYQTIDKLD-D 241
DB 271 NKKIIVSHAYCPNFPWAGLAMADGGTNAMWINDSKQSEYTFWFMNDYNTKYSRGIPVII 330
QY 242 PN-LIATVHYGFWPFS-VNIA-GYTR-FE-EDSKREIET-F-DRVHTFVARGIPVL 294
DB 331 GEGGAV--DKNN--LK-T-RVEYMSYVAQAKGILCILWDNN-NFS 371
QY 295 GEFGLGFKDKHTGVIQOGEKLFKEVLIHLNERNDRITHMLWDNGQHPN 342
```

RESULT 13

```
ENTRY
TITLE S16559 #type complete
ALTERNATE_NAMES cellulase (EC 3.2.1.4) - Ruminococcus flavefaciens
ORGANISM #formal_name Ruminococcus flavefaciens
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
ACCESSIONS S16559
REFERENCE S16559
#authors Cunningham, C.; McPherson, C.A.; Martin, J.; Harris, W.J.; Flint, H.J.
#journal Mol. Gen. Genet. (1991) 228:320-323
#title Sequence of a cellulase gene from the rumen anaerobe Ruminococcus flavefaciens 17.
#cross-references MUID:91360084
#accession S16559
#status preliminary
#molecule_type DNA
#residues 1-455 #label CUN
#cross-references GB:S55178; NID:g234871; PID:g234872
KEYWORDS glycosidase; hydrolase
SUMMARY #length 455 #molecular-weight 52420 #checksum 7708

Query Match 11.18; Score 470; DB 2; Length 455;
Best Local Similarity 29.48; Pred. No. 1.52e-59;
Matches 78; Conservative 68; Mismatches 102; Indels 17; Gaps 16;

Db 74 LDSAGLETGCGCPSEASQELFADAKAGFNTVRIPPTWFOHLDEN-DN-IDPAWMARVH 131
QY 52 FDAVGQD-ETANGNPRVTRELIADGYSIRIPVTWENRIGGAPDPIDPQFLNRVD 110
DB 132 QYVDYAYNIGLYIINLHHEQWNRADATAYDDINPRLKMLTQIATEFKDYDQHLIF 191
QY 111 EVQWALEEDLYVMINLHD-SWLWIYEMEHNYNGVMARYSLWEQLSNHFKDYPYTKLMF 169
DB 192 ECMNEPRAMPTPWNWATPVEERDVINRLEANEVEIRGMDGPYAKTRLMLPGYVASS 251
QY 170 ESNVEPK-F-SQ-NW-GEIRENHALLDLDLNTVEFVIRSGGQ-NDIRPLVLPMTETAT 224
DB 252 DKTFNLQIVLP-EN-DD-FVAVSIHAYTPYNTMTTEGAYHDTTKESNDLAYNLQ 308
QY 225 SQPLNNLYQTIDKLDNPLJATVHYGFWPFSVNI-A--G-Y-TRFEEDSKREIETFD 279
DB 309 NFRDMFINKDIPVVGEMGTSDFCN 333
QY 280 RVHTFVARGIPVVLGEFGLGFDK 304
```

RESULT 14

```
ENTRY
TITLE I40234 #type complete
ALTERNATE_NAMES carboxymethylcellulase - Prevotella ruminicola
ORGANISM #formal_name Prevotella ruminicola
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
ACCESSIONS I40234
REFERENCE I40234
#authors Matsushita, O.; Russell, J.B.; Wilson, D.B.
#journal J. Bacteriol. (1990) 172:3620-3630
#title Cloning and sequencing of Bacteroides ruminicola B-1-4 endoglucanase gene.
#cross-references MUID:90299778
#accession I40234
#status preliminary; translated from GB/EMBL/DBDJB
#residues 1-363 #label RES
#cross-references GB:M38216; NID:g143940; PID:g143941
SUMMARY #length 363 #molecular-weight 40526 #checksum 7557
```

```
Query Match 10.68; Score 447; DB 2; Length 363;
Best Local Similarity 29.28; Pred. No. 2.27e-55;
Matches 79; Conservative 69; Mismatches 108; Indels 15; Gaps 11;
```

Db	137	IWEQIAQRFKGSKNLYFEILNEP---H--GNITDSQ---INDMKRIILNIIRKT---NP	185
Qy	152	LWEQLSHHFDYPTKLAFESVNEPKFQSNWGEIRENHALLDDLNTVFEIVRQSGQND	211
Db	186	TRN-VI--IGAGYWSY-NLSQ--LEIPDDPNLIATFHYDDYSPFTHQWG--TWGKNDM	239
Qy	212	IRPLVPLTMTETASQPLLNLNYQTIDKDDPNLIATVHYGFWPFSVNIAGYRFEEDSK	271
Db	240	DAIMVFNHYK-KWSDKNNPVLYGEYGVMGHSDRSAY	277
Qy	272	REIETEDRVHHTFVARG-IPVVLGEFLGLGFKDKGTGI	309

Search completed: Fri Jun 4 10:05:29 1999
Job time : 76 secs.

```

RESULT      15
ENTRY
TITLE      cellulase (EC 3.2.1.4) CCD precursor - Clostridium sp.
ALTERNATE_NAMES
ORGANISM   #formal_name Clostridium sp.
DATE       31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
              22-Nov-1996
ACCESSIONS
REFERENCE  JQ1229
            #type complete
            #authors  Shima, S.; Igarashi, Y.; Kodama, T.
            #journal  Gene (1991) 104:33-38
            #title    Nucleotide sequence analysis of the endoglucanase-encoding
                    gene, celCCD, of Clostridium cellulolyticum.
            #cross-references MUID:92009193
            #accession JQ1229
            #molecule_type DNA
            ##residues 1-584 ##label SHI
            ##experimental_source ATCC 35319
            ##note     the authors translated the codon ACG for residue 44 as
                    Phe and TTT for 45 as Thr

```

```

GENETICS
#gene
FUNCTION
#description
celCCD
hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1,3-linkages
cellulose degradation
#superfamily Clostridium cellulase repeat homology
glycosidase; hydrolase; polysaccharide degradation
#domain signal sequence #status predicted #label SIG\
#product cellulase CCD #status predicted #label CEL\
#region proline/threonine-rich\
#domain Clostridium cellulase repeat homology #label
CCR1\
#domain Clostridium cellulase repeat homology #label
CCR2
#length 584 #molecular-weight 66061 #checksum 2934
SUMMARY

```

[illegible]